

LearnIt::R plateforme pédagogique pour l'apprentissage de R

Philippe Grosjean & Guyliann Engels

2024-06-01

Contents

Préambule	7
1 Introduction	11
1.1 Retour d'expérience	11
1.2 Apprentissage	16
1.3 Innovation pédagogique	19
1.4 Organisation de la matière	35
2 Organisation générale	41
2.1 LearnIt::R GitHub	41
2.2 LearnIt::R LRS	43
2.3 Outils supplémentaires	45
2.4 RGPD et droit d'auteur	45
3 R & RStudio	47
I LearnIt::R GitHub	51
4 Git & GitHub	53
4.1 Installation Git et cie	53
4.2 BioDataScience-Course	53
4.3 GitHub Classroom	54
4.4 Astuces Git et GitHub	55
5 Bookdown & learntitdown	57
5.1 Création d'un bookdown/learnitdown	58
5.2 Intégration dans Wordpress	63
5.3 Contenu H5P dans bookdown	67

6 Tutoriels learnrs	71
7 Applications interactive Shiny	73
8 Exercices H5P	75
8.1 Installation de H5P	75
8.2 Enregistrement H5P xAPI	78
8.3 Gestion des contenus H5P	85
8.4 Quelques contenus H5P utiles	85
9 Projets GitHub Classroom	91
II LearnIt::R LRS	93
10 Learning Record Store & MongoDB	95
10.1 Notre “LRS”	95
10.2 Learning Analytics Platform	99
10.3 Mise à jour de MongoDB	99
11 Wordpress	101
11.1 Authentification des étudiants	102
11.2 Installation Wordpress	106
11.3 Configuration serveur et git	109
11.4 Agencement du site	113
11.5 Variables contextuelles	116
11.6 Redirection d’URL d’iframe bookdown	118
12 Moodle	119
12.1 Gestion des cours	119
12.2 Badges	119
12.3 Notes	120
13 Posit Connect	121
13.1 Installation de sdd.umons.ac.be	121
14 Rapport de progression	137
14.1 Présentation générale	139
14.2 Les données	139

CONTENTS	5
15 Dashboard	147
III More tools	149
16 Vidéos	151
17 Wooclap	153
A Installation	155
B Maintenance	157
B.1 Maintenance régulière	157
B.2 Nettoyage à la fin de chaque cours	157
C Anonymisation des données	159
Références	161

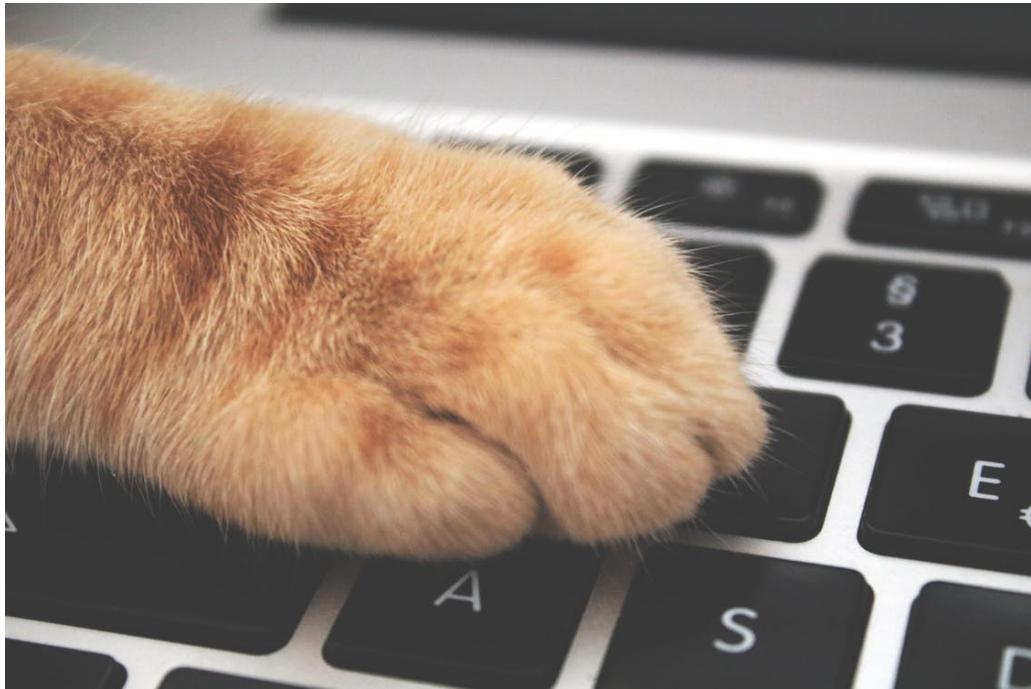
Préambule



Cet ouvrage est en cours d'écriture. Plusieurs chapitres doivent encore être écrits ou remaniés. Une version anglaise doit également être rédigée. Cependant, les informations qu'il contient sont déjà utilisables pour mettre en place certains des outils documentés ici. N'hésitez pas à nous contacter (Philippe Grosjean ou Guyliann Engels) pour plus d'information ou pour discuter de l'intégration de tel ou tel outil dans vos propres cours.



Cet ouvrage détaille les outils qui composent la **plateforme pédagogique LearnIt::R** spécifique pour l'apprentissage de R, des statistiques et de la science des données en général. Il est dérivé de notes relatives au développement, à la maintenance et à la création de contenu sur notre plateforme autour du package R `{learnitdown}` que nous avons créé à l'Université de Mons en Belgique pour nos cours de Science des Données Biologiques. Désolé, donc, si vous y lirez de temps en temps des informations qui ne sont utiles **que** dans le cadre spécifiquement de ces cours. Notre objectif, en cours de réalisation, est de migrer progressivement cet ouvrage et les outils qu'il décrit vers un usage plus large pour permettre de déployer une plateforme pédagogique similaire dans un autre contexte...



Le matériel dans cet ouvrage est distribué sous licence CC BY-NC-SA 4.0.

Détails concernant le système utilisé pour compiler ce bookdown...

```
sessioninfo::session_info()
```

Information système

```
# - Session info -----
# setting value
# version R version 4.2.3 (2023-03-15)
# os      macOS 14.4.1
# system  aarch64, darwin20
# ui      X11
# language (EN)
# collate en_US.UTF-8
# ctype   en_US.UTF-8
# tz      Europe/Brussels
# date    2024-06-01
# pandoc  3.1.11 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/aarch64/
#
# - Packages -----
# package     * version date (UTC) lib source
# bookdown    0.33    2023-03-06 [2] RSPM (R 4.2.3)
# cli         3.6.1   2023-03-23 [2] RSPM (R 4.2.0)
# digest      0.6.31  2022-12-11 [2] RSPM (R 4.2.0)
```

```
# evaluate      0.20   2023-01-17 [2] RSPM (R 4.2.3)
# fastmap      1.1.1   2023-02-24 [2] RSPM (R 4.2.0)
# htmltools     0.5.5   2023-03-23 [2] RSPM (R 4.2.3)
# knitr        1.42    2023-01-25 [2] RSPM (R 4.2.3)
# rlang         1.1.1   2024-01-06 [2] Github (r-lib/rlang@564f176)
# rmarkdown     2.21    2023-03-26 [2] RSPM (R 4.2.3)
# rstudioapi    0.14    2022-08-22 [2] RSPM (R 4.2.0)
# sessioninfo   1.2.2   2021-12-06 [2] RSPM (R 4.2.0)
# xfun          0.38    2023-03-24 [2] RSPM (R 4.2.3)
# yaml          2.3.7   2023-01-23 [2] RSPM (R 4.2.0)
#
# [1] /Users/phgrosjean/Library/R/arm64/4.2/library
# [2] /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/library
#
# -----
```

Chapter 1

Introduction

Chaque chapitre couvre un outil, avec le chapitre 2 qui présente la façon dont les différents outils sont assemblés pour implémenter l'ensemble de la plateforme pédagogique LearnIt::R et le chapitre 3 qui détaille les différentes façons de configurer un environnement logiciel pour les étudiants et les enseignants. La suite de cette introduction explique les choix effectués pour la plateforme et son utilisation dans le cadre du cours de science des données biologiques à l'Université de Mons. Ces choix visent un but prioritaire : l'apprentissage **progressif** et efficace de R et de la science des données.

Si vous recherchez un outil en particuliers, vous pouvez passer ce chapitre sans problème. Si vous êtes intéressé • e par les justification pédagogiques qui sous-tendent la plateforme, vous trouverez des éléments de réflexion dans les section du présent chapitre.

TODO: retravailler tout cela pour le rendre plus concis et mettre les éléments importants relatifs à l'utilisation de la plateforme LearnIt::R en avant.

- La démarche personnelle des auteurs qui a mené à cette plateforme est d'abord expliquée dans la section 1.1.
- Les éléments importants relatifs à l'apprentissage, tels que les psychologues expérimentalistes et les neuroscientifiques nous les expliquent sont détaillés dans la section 1.2.
- Quelques aspects qui nous paraissent importants de pédagogie et de psychopédagogie sont ensuite résumés dans la section 1.3.
- Enfin, la scénarisation, le timing et le mode d'évaluation adoptés sont expliqués à la section 1.4 par rapport à nos objectifs pédagogiques. Les outils nécessaires sont cité dans ce contexte.

1.1 Retour d'expérience

La plateforme pédagogique LearnIt::R n'a pas été conçue hors contexte. Au contraire, c'est le fruit d'une expérience d'enseignement des statistiques depuis plus de 20 ans à des étudiants en formation universitaire en biologie. À l'Université de Mons (UMONS dans le Hainaut, en Belgique), nous enseignons l'analyse statistique de données biologiques depuis la seconde année (Licence en France, Bachelier en Belgique) jusqu'à la dernière

année du Master. Nous abordons donc tous les aspects depuis les base jusqu'aux notions les plus avancées.



Les outils LearnIt::R sont utilisables dans un contexte bien plus vaste que celui décrit ici : formations diverses en présentiel (formation d'entreprise, en haute école...) ou en distanciel (MOOC, tutoriels en ligne...). Le contexte universitaire dans lequel ils ont été développés permet, néanmoins, de mieux comprendre l'architecture de la plateforme complète.

De même, cette plateforme est utilisable pour enseigner bien d'autres matières que la science des données à des biologistes, bien entendu. Sa grande force est tout de même de faire la part belle au langage R. Mais elle est parfaitement adaptable pour enseigner, par exemple, Python, Julia, ou même, une matière qui ne nécessite pas l'apprentissage d'un langage de programmation.

1.1.1 Cursus universitaire belge

Dans cet ouvrage, nous utiliserons la dénomination belge pour les cursus universitaires. Voici une présentation succincte pour les lecteurs des autres pays ou ceux qui ne sont pas familiers avec les études universitaires. Il est important de connaître cette terminologie pour la suite de cette section.

Un parcours universitaire belge classique suit les prescriptions européennes. Il est composé d'un **premier cycle** de trois années de **Bachelier**, suivi de deux ans de **Master** en **second cycle**, éventuellement prolongé d'un **troisième cycle** sous forme d'une thèse de **doctorat**.

Une année universitaire est appelée **année académique**. Elle commence début septembre avec un **premier quadrimestre** (13 semaines dont 12 de cours et des examens en janvier). Le **second quadrimestre** est également composé de 13 semaines d'activités dont 12 semaines de cours qui se terminent vers la mi-mai. Une à deux semaines de "**blocus**" où les étudiants se préparent pour les examens, précède la **première session d'examens** de juin. Enfin, une seconde période de blocus précède la **seconde session d'examens** d'août qui clôture l'année académique. Comme les années académiques sont à cheval sur deux années civiles, on parlera de l'année académique 2023-2024, par exemple, pour celle débutant en septembre 2023 et se terminant fin août 2024.

Une année académique est composée d'une série de matières qui peuvent être soit obligatoires, soit à option. Ces matières sont constitués d'exposés plutôt théoriques en amphithéâtres réalisés par les **professeurs** (les **cours ex cathedra**, ou **cours théoriques**, dans d'autres pays on parlera de cours magistraux), accompagnés éventuellement de séances plus pratiques d'**exercices** ou de **travaux pratiques**, généralement dispensés par un ou plusieurs **assistants**. En Belgique, les assistants "classiques" sont diplômés de Master et prennent leurs enseignements à mi-temps. Ils réalisent en parallèle un travail de recherche en vue de présenter une thèse de doctorat pendant l'autre mi-temps, et ce, sur une durée de six ans.

Chaque matière est reprise comme une **Unité d'Enseignement** (UE), décomposée en une ou plusieurs **Activités d'Apprentissage** (AA), elles-mêmes constituées d'un certain nombre d'heures de cours, de séances d'exercices, de travaux pratique ou autres (par exemple, des excursions ou des stages). En fonction du volume horaire de travail total estimé pour l'étudiant, un nombre de **crédits**, encore appelés **ECTS** (pour "*European*

Credit Transfer and Accumulation System") est attribué à chaque UE et chaque AA. Un crédit ECTS correspond à une charge de travail de 25h à 30h pour l'étudiant. C'est une définition imposée à tous les états membres de la Communauté Européenne.

Une année académique belge est constituée de matières totalisant **60 crédits ECTS**. Le Bachelier compte donc 180 ECTS au total et le Master en deux ans 120 ECTS. Comme exemple, voyez ici le programme de cours du Bachelier en Biologie à l'UMONS durant l'année académique 2023-2024. Vous pouvez y repérer l'UE de "science des données biologiques I : visualisation et inférence" en **bloc 2** (c'est à dire, en seconde année du Bachelier), ainsi que l'UE de "science des données II : analyse et modélisation" en bloc 3 dont il sera question ci-dessous. Vous noterez que ces deux UE *n'ont aucune heure de cours théoriques* (colonne HT). Toutes les heures en présentiel (70h et 60h, respectivement) sont des HTPPE ou heures de travaux pratiques encadrés. Nous y reviendrons. Ces UE correspondent à, respectivement 6 et 5 crédits ECTS et se répartissent sur les deux quadrimestres. Bien que ce ne soit pas une règle absolue, les différentes UE sont plutôt enseignées en parallèle tout au long d'un ou des deux quadrimestres, avec des séances de 2h ou de 4h disséminées à l'horaire tout au long des 12 ou des 24 semaines de cours.

1.1.2 Evolution de notre approche pédagogique

La formation en science des données biologiques telle qu'enseignée en 2023-2024 n'a plus rien à voir avec la formation en biostatistique de 2003-2004. En voici l'évolution, et surtout, les raisons qui en ont motivé les changements :

- Cours théoriques et séances d'exercices à volume horaire égal. Les cours de biostatistiques étaient constitués initialement de cours ex cathedra avec support de type PowerPoint pour autant d'heures que d'exercices. Ces derniers étaient réalisés sur ordinateurs avec le logiciel R. Les étudiants soumettaient un rapport de leurs analyses au format Word. Ce rapport était coté et la partie théorique était évaluée indépendamment lors d'un examen oral en session. Constatation : l'objectif pour des biologistes étant plus la capacité à réaliser des analyses statistiques de données biologiques que de connaissances théoriques en statistiques, nous avions la volonté de privilégier les séances d'exercices.
- Même mode d'enseignement, mais volume horaire de 1 pour 2 entre cours théoriques et séances d'exercices. Il en résulte une partie théorique raccourcie, avec notamment moins de développements mathématiques et plus de données biologiques réelles analysées avec R par les étudiants sur les ordinateurs. Constatation : les étudiants restent réfractaires à la partie théorique, mais s'investissent plus dans les séances pratiques. Ils ne s'intéressent à la théorie qu'après avoir constaté l'intérêt dans le cadre du travail d'un biologiste sur base des cas concrets traités durant les séances d'exercices. C'est le problème de l'œuf et de la poule. Il faut une base théorique minimale pour aborder les exercices, mais nos étudiants ne s'intéressent à la théorie qu'après avoir fait les exercices.
- Même mode d'enseignement et volume horaire que ci-dessus, mais répartition différente dans le temps (les séances de cours et d'exercices étaient totalement découplées jusqu'ici). Une séance "composite" est constituée d'un partie théorique avec démonstration immédiate sur un cas concret. Les étudiant effectuent les exercices directement par la suite. Tant la démonstration plus théorique que l'application pratique se font dans la salle informatique. Les étudiants peuvent donc exécuter les scripts R de la démonstration sur leurs ordinateurs. Les parties "cours théoriques"

sont donc moins passives de la part des étudiants. Constatation : intégration des deux séances est bénéfique, mais le cours théorique en amont reste une partie où les étudiants sont les moins attentifs.

- Suite à notre intérêt pour les techniques pédagogiques alternatives, et notamment les classes inversées et les enseignement hybrides où une partie du travail est réalisé par l'étudiant à domicile avant les séances, nous sommes repartis d'une page blanche. La même année, nous avons décidé d'élargir la matière au delà des biostatistiques en transformant les anciennes UE de biostatistiques en UE de science des données où nous abordons aussi plus en détails le stockage et la manipulation des données, ainsi que la présentation des résultats et l'esprit critique statistique. Une nouvelle UE est ajoutée en Master 2 et les volumes horaires et ECTS des autres UE sont augmentés pour atteindre cet objectif. Il n'y a plus du tout d'heures de cours théoriques. Toutes les séances en présentiel deviennent des heures d'exercices encadrés, ainsi que de coaching pour répondre à leurs questions sur la matière théorique ou la réalisation de leurs analyses. La présence à ces séances devient obligatoire. La matière est divisée en 30 modules dispensés une semaine sur deux pour assurer l'apprentissage progressif et sur la durée. Le contenu est réécrit en bookdown pour être accessible en ligne (<https://wp.sciviews.org>). Les étudiants doivent lire le chapitre correspondant et réaliser des tutoriel learnr avant la première séance du module. Les exercices se font maintenant avec R, RStudio et GitHub pour gérer les dépôts des projets des étudiants. Le format des rapports passe à R Markdown. Il n'y a plus d'examen théorique, uniquement des projets cotés. Constatation : l'approche progressive en différents modules et l'évaluation continue sont bien utiles. Les tutoriels learnr sont très appréciés. Mais seule une petite fraction des étudiants prépare la matière à l'avance. Les autres réalisent les tutoriels learnrs ou lisent le cours en ligne en classe. Ils ne passent pas assez de temps à poser des questions et à réaliser leurs analyses en séances. Ensuite, ils bâclent généralement leurs rapports pour les finir chez eux avant la deadline. Une enquête auprès des étudiants montre que le cours en ligne n'est pas suffisant pour apprêter efficacement la matière chez eux avant les séances.
- Les séances sont mieux scénarisées et le timing défini plus clairement. Des exercices sont rajoutés dans le cours en ligne (H5P et application interactives Shiny) pour vérifier l'auto-apprentissage des étudiants et leur permettre de déterminer s'ils ont bien compris les différents concepts. Les tutoriels learnrs sont complétés de suggestions à l'aide du package gradethis. Une base de données MongoDB est mise en place comme LRS (“learning record store”) de fortune et du code est créé pour collecter les traces de l'activité des étudiants dans les exercices H5P, Shiny et learnr (c'est le point de départ du package learnitdown de la plateforme LearnIt::R). Des points sont attribués (5 à 10% de la note finale selon les années) à la résolution de ces exercices. Constatation : le cours en ligne agrémenté d'exercices d'auto-évaluation avec suggestions automatisées pour corriger les erreurs, notamment dans le code R des tutoriels learnrs est une formule bien plus efficace pour l'approche classe inversée. Les points attribués à la résolution de ces exercices incite fortement les étudiants à les réaliser avec un taux de participation aux exercices en augmentation. La préparation est bien meilleure, mais encore perfectible. Un trop grand nombre d'étudiants ne préparent toujours pas suffisamment à l'avance.
- Un rapport de progression dans les exercices destiné aux étudiants et une section récapitulative des différents exercices à la fin de chaque chapitre du cours en ligne sont rajoutés (voir par exemple ici). Le rapport de progression individuel s'appuie sur les données collectées dans le LRS et est écrit en R et en Shiny. Il deviendra

plus tard le package learnitprogress de la plateforme LearnIt::R. Des fonctions sont rajoutées dans le package learnitdown pour automatiser la création de la section récapitulative et encore mieux intégrer les exercices H5P, Shiny et learnr dans le cours. Constatation : le taux de participation aux exercices atteint des niveaux inespérés (plus de 90% et certaines années jusqu'à 98% des exercices réalisés par les étudiants en moyenne). Des difficultés deviennent plus apparentes dans la transition entre théorie et pratique maintenant que les outils ad hoc sont en place pour la partie plus théorique.

- Mise en place de projets GitHub individuels guidés (voir un exemple [ici](#) et [ici](#)). Des projets en groupes de deux à quatre étudiants *non guidés* forment toujours le niveau de difficulté ultime. Les projets individuels guidés ont de nombreuses instructions et une batterie de tests basés sur le package testthat qui permet aux étudiants de vérifier par eux-mêmes s'il ont obtenu la bonne réponse et de leur faire des suggestions dans le cas contraire (feedback immédiat et détaillé). L'interprétation des résultats se fait en sélectionnant les bonnes phrases dans une section à choix multiple dans le document R Markdown ou Quarto. Cela leur montre comment rédiger leur interprétation et ce à quoi il faut faire attention à ce niveau-là (ils devront rédiger leur interprétation par eux-mêmes ensuite dans les projets de groupe). Les projets sont corrigés selon une grille critériée remplie de manière semi-automatique. Du code R est écrit pour gérer tout cela. Ce code se concrétise finalement par le package learnitgrid de la plateforme LearnIt::R. Constatation : la transition entre théorie et pratique est plus douce. Nous avons maintenant quatre niveaux de difficulté croissante dans les exercices (H5P et Shiny dans le cours = niveau 1, tutoriels learnrs = niveau 2, projets GitHub individuels cadrés = niveau 3, projets GitHub en groupe libres = niveau 4). L'apprentissage est plus progressif et plus homogène au niveau de la classe. Des différences temporelles importantes s'observent cependant toujours au sein de la classe entre les étudiants qui respectent le timing proposé et ceux qui font tout à la dernière minute avant les deadlines. Cela pose problème pour la bonne progression dans les travaux de groupes.
- Un outil d'analyse des traces d'apprentissages des étudiants telles que récoltées dans notre LRS est écrit en Shiny (il deviendra le package learnitdashboard de la plateforme LearnIt::R). Il nous permettra de repérer plus facilement les difficultés restantes et d'y remédier à l'avenir. La scénarisation des séances est encore revue en séparant clairement le travail dans les projets individuels durant une première séance et les travaux de groupe lors d'une seconde séance. Une courte interrogation écrite sur la matière du module, y compris le projet individuel en début de seconde séance vient forcer les étudiants à mieux respecter le timing. Ces dernières modifications feront l'objet d'une réévaluation l'année prochaine. Une pondération plus judicieuse entre exercices, les projets individuels, les projets de groupe et les interrogations permettra une notation la plus adéquate possible de l'acquis de chaque étudiant.

Un aspect important à noter ici, c'est que notre approche pédagogique n'a pas été élaborée d'un seul coup et sur des bases uniquement théoriques de pédagogie. Nous avons observé et amélioré année après année notre méthode. Nous avons inclus des outils pédagogiques et logiciels nouveaux pour nous (ou les avons programmés nous-mêmes s'il n'existaient pas) *très progressivement* sur un intervalle de vingt années, et surtout durant les six dernières. Nous ne retenons au final que ce qui permet une amélioration significative de l'apprentissage. Faites-en de même : incluez vos changements de manière très progressive en vous assurant de bien maîtriser les nouveaux outils ou les nouvelles techniques *avant* de les inclure dans vos approches pédagogiques. Les étudiants ne sont pas des cobaye !

Méfiez-vous de ce que vous conseillent les techno-pédagogues et n'adoptez que les techniques et les outils qui *vous* donnent de bons résultats dans le contexte particuliers de *vos cours*.

Ceci dit, si une approche empirique de vos enseignements est utile comme nous, nous l'avons privilégiée, la science a tout de même bien progressé pour décrypter les mécanismes de l'apprentissage. La section suivante reprend diverses notions issues de tels travaux qu'il est utile de connaître pour améliorer votre façon d'enseigner.

1.2 Apprentissage

Les neuroscientifiques et les psychologues ont fait beaucoup de progrès ces derniers temps pour comprendre comment le cerveau humain apprend. Ces découvertes ont des retombées importantes au niveau de la pédagogie. Il est donc utile de s'y attarder un peu dans le but de dégager les stratégies d'apprentissage gagnantes (et celles qui sont perdantes).

Le schéma suivant résume comment l'apprentissage se fait dans notre cerveau.

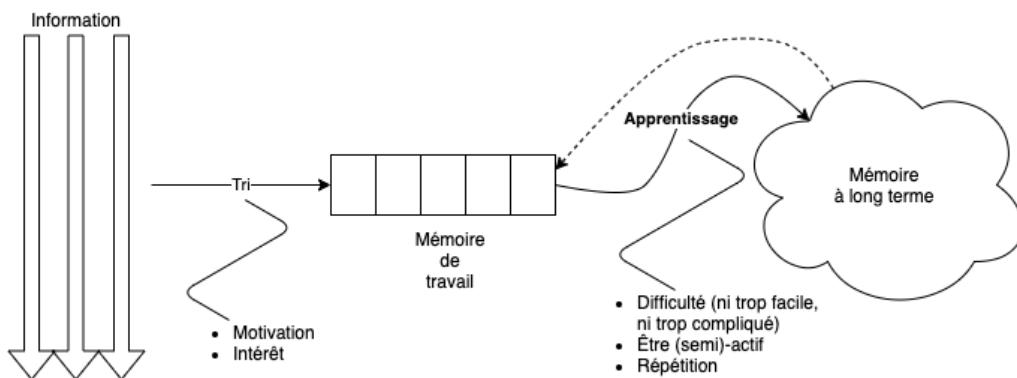


Figure 1.1: Mécanismes menant à l'apprentissage.

- Nos organes des sens sont noyés en permanence dans un flux d'informations. Il est impossible de toutes les traiter avec la même attention. Ainsi, un tri est réalisé très tôt et de manière (presque) inconsciente. Tout ce qui n'est pas conservé est bien entendu perdu. Les critères qui motivent le tri de l'information considérée comme utile sont la **motivation** et l'**intérêt**. Par exemple, donner une longue explication à des étudiants *avant* d'avoir suscité leur motivation ou leur intérêt est contre-productif, car cette information sera filtrée et éliminée très tôt sans même être traitée ! C'est le coup classique : s'arrêter en pleine explication et demander à un étudiant au hasard, ou à un étudiant manifestement peu concentré ce qu'on vient de dire... et il est incapable de le répéter ! Donc, arriver à cette situation *est en partie de la faute du professeur*. Il faut qu'il prenne soin de susciter l'intérêt et/ou la motivation *avant* de donner des explications compliquées. Cela peut-être par une phrase du genre : **attention, ce que je vais vous expliquer maintenant vous servira pour faire X ou Y plus tard... ou soyez bien attentifs, la procédure suivante est cruciale pour la réussite de l'exercice....** Cela est également valable dans le matériel pédagogique où il faut bien préciser les objectifs

en début de section (et que ces objectifs fassent écho -motivation ou intérêt- dans la tête des étudiants).



Ce qui n'est pas perçu comme "utile" par l'étudiant au moment où cela est formulé est *immédiatement* oublié... C'est ce processus de tri de l'information pertinente qui en est responsable !

- Une fois le filtre du tri franchi, l'information à traiter arrive dans la mémoire de travail à court terme. Cette mémoire permet un traitement rapide de l'information, **mais elle est limitée !** Selon les individus et/ou les circonstances, elle ne dépasse pas l'équivalent de 7-9 "cases" maximum. Il vaut mieux même considérer **cinq cases** disponibles. Chaque case est capable d'accueillir une information à traiter. Cela signifie qu'il faut éviter d'amener trop de difficultés à la fois. Si le problème est complexe et nécessite l'assimilation de nombreuses nouvelles notions, il faut alors découper le problème en plusieurs étapes plus simples. Il faut aussi faire attention qu'un certain nombre de ces cases peuvent se remplir par des éléments parasites. Par exemple, un étudiant qui doit installer un logiciel sur son ordinateur, mais qui se bat avec l'écran de login parce qu'il ne l'entre pas correctement, ou parce qu'il n'arrive pas à retrouver son mot de passe, va concentrer son attention et une partie de son "espace de travail" à résoudre ce problème de login. Ensuite, l'espace de travail utilisé ne sera pas directement réutilisable pour la tâche désirée (installation du logiciel). L'étudiant va se sentir débordé par les informations et en conclura qu'il est nul en informatique. Si le même étudiant avait pu se logger directement sans problèmes, il n'éprouverait peut-être pas de telles difficultés.



Il faut toujours garder à l'esprit la taille très limitée de l'espace de travail dans notre cerveau, et ne jamais amener trop de notions nouvelles et/ou difficiles d'un coup. Il faut aussi apprendre à identifier les situations où les étudiants sont saturés d'information. Dans les deux cas, il faut décomposer alors le problème en étapes successives plus simples et attaquer étape par étape.

- Si l'information arrive dans l'espace de travail du cerveau de l'étudiant, ce n'est pas encore gagné pour autant. Il faut en effet bien évidemment qu'il soit capable de traiter le problème. S'il est dans une impasse, tout s'arrête-là ! **Donc, il faut adapter le niveau de difficulté aux connaissances et au potentiel de l'étudiant.** La difficulté ici, c'est qu'une classe est hétérogène. Donc, le niveau de difficulté acceptable n'est pas le même pour tous. Varier le matériel pédagogique et proposer des exercices de niveaux de difficulté différents permet de résoudre ce casse-tête, mais c'est aussi crucial pour l'apprentissage pour une autre raison (voir ci-dessous).
- Une fois l'information traitée, elle ne mène pas nécessairement à un *apprentissage* pour autant. Notre cerveau a évolué au cours du temps pour économiser ses ressources. Ainsi, ce qui est trop facile à traiter ne mérite pas d'être appris. En effet, dans ce cas on est capable de refaire le même raisonnement facilement si la situation se présente. Donc, notre cerveau va éviter de *mémoriser* des éléments qui paraissent trop simples (le verbe "paraître" est important, car il fait intervenir une sensation.... qui peut être erronée et faire alors perdre la possibilité d'apprendre un concept important).

- Ainsi une des conditions nécessaires pour que le concept étudié soit transféré dans la mémoire à long terme est qu'il soit raisonnablement difficile (sinon, il ne sera pas digne d'être mémorisé), mais pas trop (si l'étudiant ne peut pas résoudre le problème auquel il fait face, il ne peut rien mémoriser d'utile bien évidemment).



Adapter le niveau de difficulté des activités proposées est à la fois complexe (à cause de l'hétérogénéité des classes) et **indispensable** pour que le mécanisme d'apprentissage s'enclenche. Une façon d'arriver à un niveau de difficulté adéquat, ni trop simple, ni trop difficile, consiste à reporter partiellement cette tâche sur l'étudiant lui-même. Si l'étudiant devient (partiellement) responsable de son processus d'apprentissage, il pourra ajuster partiellement de lui-même le niveau de difficulté (choix des exercices, de leur ordre, de leur timing, avoir recours à l'entraide...)

Une approche complémentaire indispensable consiste à varier le niveau de difficulté des exercices proposés autour d'un même concept. Les étudiants les moins performants apprendront avec les exercices les plus simples, même s'ils bloqueront probablement sur les cas les plus complexes. D'un autre côté, le génie survolera ces exercices simples sans rien apprendre, mais s'attardera de manière utile sur les exercices plus difficiles, et c'est là qu'il apprendra le mieux.

- Une autre condition indispensable pour que le mécanisme d'apprentissage s'enclenche, c'est d'être **actif**. Or, écouter un professeur déblatérer sa matière devant un grand auditoire, visionner une vidéo, et parfois même, lire un long texte continu, ce n'est **pas** être actif. L'information "glisse". On dit quelle "entre par une oreille et ressort par l'autre". Et c'est vraiment pratiquement comme cela que cela se passe : le mécanisme d'apprentissage n'est pas activé. Par contre si l'étudiant fait quelque chose pendant ce temps-là, quand il est actif, alors il apprend. Le paradoxe, c'est que l'activité ne doit pas nécessairement être liée à la matière à apprendre. Par exemple, un parent qui cherche à aider son enfant pour apprendre ses devoirs en répétant avec lui sera énervé si l'enfant joue à un jeu vidéo en même temps... Pourtant, c'est l'enfant qui a raison : il est actif (sur son jeu), donc il apprend (sa matière) ! Ce n'est pas logique, mais ainsi fonctionne notre cerveau.



La pédagogie active vise à activer ce mécanisme d'apprentissage par l'activité autant qu'elle reporte la responsabilité de choix de difficulté et de mécanisme d'apprentissage sur l'étudiant lui-même qui doit développer une certaine autonomie pour mettre en place les conditions qui font que, pour lui, l'apprentissage sera optimal. Les enseignants doivent encourager cela. Certains psychopédagogues modernes sont très critiques sur l'enseignement via les grands cours ex cathedras en université. Certains disent que c'est la meilleure méthode pour que la matière passe directement du PowerPoint du prof à la feuille de notes de l'étudiant sans passer par le cerveau du professeur, ni par celui de l'étudiant. **Le mécanisme d'apprentissage est pratiquement inexistant ici, et c'est plus tard, en remettant de l'ordre dans ses notes de manière active que l'étudiant apprendra véritablement la matière.**

- Un dernier point important pour l'apprentissage est la **répétition**. Pour qu'un concept s'imprime durablement dans le cerveau, il faut que les connexions synaptiques soient activées et réactivées *plusieurs fois*. Ceci est un concept connu depuis longtemps en pédagogie, mais qui est parfois mal appliqué. Car si cette répétition est monotone, elle est ennuyeuse. Et alors, la motivation et l'intérêt de l'étudiant sont en chute libre... au point que l'information répétée peut très bien ne plus passer la première barrière du tri. Dans ce cas, la répétition ne sera plus daucun effet bénéfique.



Il faut répéter les concepts pour qu'ils s'ancrent durablement dans le cerveau des apprenants. Mais cette répétition ne doit pas être monotone. Donc, il faut multiplier les méthodes de transmission du concept. Le multimédia permet de varier (texte, image, son). On peut aussi diversifier les exercices, approcher de manière ludique, inventer des challenges, etc. **Donc, cette répétition passe encore une fois par la diversité du matériel pédagogique fourni à l'étudiant autour d'un même concept.**

Un leitmotiv, on le voit très clairement ici, qui ressort de tout ceci est qu'il faut **diversifier le matériel pédagogique** autant que possible. C'est notre objectif dans la plateforme pédagogique LearnIt:RR, et c'est pour cela que nous proposons de nombreux outils différents qui ne ressemblent pas du tout à un syllabus qui contient du texte et des images statiques, accompagné de PowerPoints qui contiennent aussi du texte et des images statiques (ou pire, des "animations" qui n'apportent rien à l'apprentissage... la fameuse mort par PowerPoint).

Il n'y a pas que les psychologues et les spécialistes du cerveau qui ont fait des progrès dans la compréhension des mécanismes d'apprentissage. Les techno-pédagogues ont aussi développé des nouvelles techniques et d'autres outils pour mieux apprendre. C'est l'innovation pédagogique dont nous allons reprendre certains éléments dans la section suivante qui nous ont paru importants dans l'élaboration de la plateforme pédagogique LearnIt:R.

1.3 Innovation pédagogique

Cette section reprend différents éléments importants qui émergent des constatations faites par les enseignants qui se sont lancés dans l'aventure ainsi que des techno-pédagogues... sans chercher forcément un fil conducteur entre ces différentes idées.

1.3.1 Différents types de classes inversées

- Marcel Lebrun de l'Université Catholique de Louvain présente les différents types de classes inversées dans la vidéo suivante :

À côté de l'approche traditionnelle cours ex cathedras qui présentent la théorie suivis de séances d'exercices ou TP qui mettent en pratique, il y a :

- la **classe inversée** “traditionnelle” (type 1 selon lui) où les étudiants apprennent la matière à la maison et font les devoirs en classe. Cela reste la théorie d'abord et l'application ensuite. De même, la matière reste imposée par l'enseignant.
- la **classe renversée** (type 2 dans sa classification). Ici, ce sont les étudiants qui élaborent le matériel pédagogique et le prof est un facilitateur. On explore le contexte d'abord pour définir quoi étudier, et on réalise une analyse et une synthèse ensuite.
- la classe inversée de type 3 mélange les deux dans un processus itératif de contextualisation -> décontextualisation -> recontextualisation.



Il n'y a pas qu'un seul type de classe inversée. On peut renverser les lieux sans toucher au processus (type 1), mais on peut aussi changer les rôles avec la classe renversée où ce sont les étudiants qui élaborent et expliquent le cours au prof (classe renversée). Bien sûr, on peut aussi panacher les deux avec d'autres parties en forme plus classique cours ex cathedras suivis d'exercices.

Les différents types de classes inversées peuvent se définir par rapport aux savoirs (qui apporte l'info, le prof ou les élèves ?), et par rapport aux rôles (qui est actif, le prof ou les élèves ?)

1.3.2 Contextualiser/décontextualiser

Dans son exposé (voir ci-dessus), Marcel Lebrun fait remarquer que l'enseignement classique expliquer *d'abord* la théorie et passe *ensuite* aux applications. La théorie est une **décontextualisation** du savoir, tandis que les problèmes et applications pratiques en sont une **contextualisation**. Il fait remarquer que cela inverse le processus naturel hors classe, où l'humain rencontre un problème (le contexte) et cherche à le résoudre en l'étudiant et en cherchant à la généraliser pour ne plus le rencontrer, ou du moins y réagir plus efficacement s'il se reproduit (décontextualisation). L'enseignement traditionnel et les classes inversées de type 1 travaillent donc dans le mauvais ordre, alors que la classe renversée part bien du contexte pour décontextualiser ensuite. Au sein du matériel pédagogique, nous pouvons décider de partir de la théorie vers les applications, ou bien, de partir d'un problème concret et de raisonner à son sujet pour aboutir à de la théorie. Il semble que cette dernière approche soit plus efficace.



L'esprit humain part d'un contexte pour ensuite généraliser (décontextualiser). L'enseignement classique, en présentant la théorie avant les applications inverse ce processus (décontextualisation avant contextualisation). De nombreux indicateurs montrent que ce n'est pas une bonne idée. Donc, l'innovation pédagogique cherche aussi à renverser l'ordre dans le flux d'apprentissage.

1.3.3 Classe renversée

L'approche sous forme de classe renversée est donc la forme la plus extrême. Jean-Charles Cailliez de l'Université de Lille présente ce concept dans la vidéo TEDx suivante :

Il ne faut pas copier son système, mais en reprendre des éléments et l'adapter au cours à la personnalité de l'enseignant, ...En tous cas, lui va très loin puisque ce sont les étudiants qui font tout (définir la matière, préparer le cours, l'exposer au professeur, préparer les questions de devoirs et d'examens). Le professeur fait les devoirs que les étudiants lui donne et offre ainsi un feedback sur la qualité du travail par cet intermédiaire. Il utilise notamment les QCM (questionnaires à choix multiple) inversés : les étudiants produisent les question, le prof répond, les étudiants corrigent. Si la question est bien posée, le professeur répond bien. Si elle est mal posée, le professeur fait semblant de répondre n'importe quoi pour bien faire comprendre que la question est mal posée. Il utilise aussi le QCM "intelligent" : il ne suffit pas de dire quelles sont les bonnes réponses, mais il faut aussi justifier pourquoi les autres sont fausses pour avoir des points. Au début du cours, chaque équipe crée un tel QCM (cinq questions). Ensuite les QCM passent à l'équipe suivante qui y répond. Enfin, les réponses sont données à une troisième équipe qui corrige (3min à chaque fois, suivi d'une petite discussion).

Cette approche permet aussi de mettre en œuvre le principe suivant : "enseigner est la meilleure façon d'apprendre." Elle est aussi mise en œuvre par Sébastien Bette à l'Université de Mons. Voir aussi les vidéos suivantes :

- Intervention de Jean-Jacques Cailliez.

- Intervention de Sébastien Bette.

Questions croisées aux deux intervenants.

1.3.3.1 Intelligence collective

François Taddei parle de la mobilisation de l'intelligence collective au service de nouvelles formes d'apprentissage ici :

Parmi les points importants de l'innovation pédagogique est l'utilisation de l'intelligence collective. Jean-Jacques Cailliez (voir ci-dessus) fait travailler les étudiants par groupes de six (considéré comme la meilleure taille de groupe). Par contre, c'est lui qui crée les groupes. Après un test d'évaluation qui lui permet de déterminer qui est fort, moyen ou faible, il mets un fort, et un faible systématiquement dans chaque groupe. Ensuite, il demande que les étudiants s'attribuent à chacun un rôle au sein du groupe pour éviter que certains ne fassent rien. Au fil du temps, les responsabilités changent, mais ceci est important pour impliquer tout le monde dès le début. Si un ou plusieurs étudiants ne font rien au sein du groupe, cela plombe l'ambiance puisqu'il y a une partie de la note qui est commune à tout le groupe et ceux qui travaillent se sentent lésés.

La mobilisation de l'intelligence collective passe par la prise de conscience qu'il y a plus dans plusieurs cerveaux que dans un seul et que le travail réalisé à plusieurs est *toujours* de meilleur qualité qu'isolé. De plus, le travail de groupe imprime une toute autre dynamique à l'apprentissage et permet de le varier.

François Taddei parle de la société apprenante comme d'un tout. Il dit que la ségrégation entre apprentissage à l'école, développement de nouveaux savoirs par le recherche et utilisation du savoir ailleurs est artificielle. L'apprentissage est partout et ne se limite pas à l'école. C'est la société apprenante.

Il considère que l'enseignement doit aussi apprendre à apprendre en individuel et en collectif, afin de pouvoir remobiliser plus tard cette capacité sur le terrain. Il faut apprendre

tout au long de sa vie car le monde est en évolution permanente. La capacité à se réinventer n'est pas facile à développer et est intimidante. Donc, si on peut la développer dans un cadre bienveillant qu'est l'école, c'est mieux. Dans ce cadre il y a moins d'enjeux, et donc, de pressions. Cela permet d'offrir le **droit à l'erreur**. C'est en faisant des erreurs et en se relevant de ces erreurs qu'on peut progresser le plus.

Concernant les MOOC ("Massive Open Online Course"), les universités et écoles ont un défi à relever. Quelle valeur ajoutée proposent-elles face aux MOOC déshumanisés ? C'est une réelle réflexion à avoir. Le défi suivant après le couplage de la recherche et de l'enseignement, c'est de faire de la recherche sur la formation afin de se réinventer en permanence toujours vers de meilleures méthodes d'apprentissage.

1.3.4 Timing et rythme

Jean-Jacques Cailliez insiste aussi fortement sur l'organisation d'une séance de cours qui doit être **variée et comporter des surprises** afin d'éviter la monotonie et l'ennui des étudiants. Même la meilleure innovation pédagogique, si elle est appliquée de manière répétitive finit par devenir ennuyeuse. Il consacre les 10 à 15 premières minutes du cours avec une forte présence où il propose à chaque fois un petit exercice différent (surprise). Ensuite seulement, les étudiants sont autonomes. A la fin, il clôture souvent par un petit questionnaire où les étudiants donnent leur avis (voir point suivant).

1.3.5 Explication, implication

L'innovation pédagogique suppose d'autres méthodes d'enseignement qui impliquent que les étudiants doivent *faire un effort pour s'y adapter*. Ils ne sont pas prêts à le faire spontanément et il faut s'attendre à des réticences voire des barrages importants de leur part. Il est donc vital d'**expliquer** et d'**impliquer**.

Avant de modifier la façon dont un enseignement est dispensé, il est vital d'en informer les étudiants. Une séance de une ou deux heures qui explique les méthodes qui seront utilisées et qui les **justifient** avant les cours proprement dits permet de préparer les étudiants aux nouvelles pratiques et aide à leurs acceptations.

Nous pouvons même aller plus loin en leur expliquant que c'est une expérience et que leur avis est crucial (et ensuite leur demander d'indiquer régulièrement deux points positifs et deux points négatifs à chacun pour les impliquer réellement dans le processus). L'implication est encore plus fortement sollicité si, après l'exposé expliquant ce qu'on compte faire, on leur laisse le **choix de le faire ou non**. On les fait voter et on ne tente l'"expérience" que si une majorité est en faveur du changement (mais il faut alors être prêt à laisser tomber).

Aussi, pour les mêmes raisons, il faut être attentif à apporter des innovations de manière graduelle. Le grand coup de pied dans la fourmilière est rarement une bonne approche en matière d'innovation pédagogique !

Quoi qu'il en soit, on ne peut pas contenter tout le monde. Il faut s'attendre en général à 10-15% de mécontents, quel que soit le choix réalisé. Les bons élèves de la classe dans le système traditionnel (les "bachotteurs") se retrouvent en général moins bien dans un système d'innovation poussée telle que la classe renversée. Du coup, ils ont tendance à critiquer négativement l'innovation et sont demandeurs pour revenir à un système plus classique.

1.3.6 Enseignement basé sur les objectifs

L'enseignement basé sur les objectifs définit *d'abord* les objectifs d'apprentissage et les savoir à acquérir, les concepts à maîtriser. Ensuite, le cours et le matériel pédagogique sont construits progressivement autour des notions clés qui auront été dégagées. Il est possible d'écrire des exercices et des questions pour mettre en pratique et (auto-)évaluer l'acquisition des concepts. Ensuite la partie théorique vient chapeauter le tout sans se perdre dans des digressions inutiles par rapport aux objectifs.



L'enseignement basé sur les objectifs consiste à définir précisément les objectifs d'abord, et ensuite à rester focalisé sur ces objectifs tout au long de l'élaboration du matériel pédagogique et du cours.

1.3.7 Attention à la formulation

On a tous tendance à utiliser des tournure comme “il suffit de ...”, “cela se fait facilement comme ceci...”. Parfois ces tournures sont utilisées pour minimiser la difficulté, ou parce que cela semble effectivement facile pour l'enseignant... qui a oublié les années d'expérience derrière lui qui ont rendu cela facile pour lui. Par contre, l'étudiant qui trouve cela difficile se sentira dévalorisé par cette formulation. Il faut donc éviter ce genre de tournure de phrase : rester neutre, et c'est éventuellement l'étudiant lui-même qui en conclura que c'est facile... ou non.

1.3.8 Effets psychologiques néfastes

Les enseignants doivent tenir compte également de la psychologie, et en particulier de tout ce qui peut accroître, ou au contraire, impacter la motivation des étudiants. Dans une présentation TEDx, Mark Rober introduit ce qu'il appelle l’“effet Super Mario”.

Le jeu Super Mario est très addictif, parce que le joueur est focalisé sur l'objectif à atteindre (terminer le jeu et sauver la princesse). Cela le détourne des nombreuses embûches qui parsèment le jeu. Au final, l'utilisateur ne fait qu'appuyer successivement sur une série de six boutons (sur la manette de jeu), et présenté autrement, on réalise à quel point la même action peut être perçue négativement. Mark présente six boutons et explique qu'il faudra les appuyer selon une séquence précise... très longue et pendant plus d'une heure pour passer un test. Vu comme cela, cela paraît infaisable et totalement rébarbatif... pourtant c'est exactement ce qu'on fait dans le jeu Super Mario qui est, au contraire, addictif.

Il transpose cela aux problèmes de la vie quotidienne. Si on se focalise sur le problème qu'on a juste en face de soi, on est vite tenté d'abandonner. Si au contraire, on se focalise sur un but lointain, mais enviable, les différentes embûches qui se présentent sur notre chemin sont plus facilement traitées. C'est donc l'état d'esprit qui est le plus important pour déterminer si on laisse tomber et perds, ou si on va de l'avant et réussi. **Il y a ici matière à réflexion profonde pour les enseignants, afin de motiver leurs étudiants.**

Mark a aussi étudié le cas d'un problème à résoudre soumis à Internet avec un grand nombre de réponses obtenues. Il permettent d'étudier la motivation vers le succès de manière statistique. Il a comparé deux sous-populations (même question, mais avec une variante subtile). Le premier groupe reçoit 200 points fictifs dès qu'il réussit, quel que soit le nombre d'essais pour y arriver. Le second groupe "perd" cinq points fictifs à chaque tentative infructueuse. Le premier groupe réussit plus souvent et avec une moyenne de 12 tentatives, par rapport à la moyenne de 5 tentatives pour le second groupe. En absolu, la pénalité est très faible et purement fictive (pas de perte d'argent ou autre conséquence quelconque dans la vie réelle). Cependant, la simple idée de "perdre" quelque chose influence profondément les utilisateurs.



Les aspects psychologiques clés autour de la motivation qui fait qu'on résoud un problème ou qu'on le laisse tomber sont cruciaux. L'effet Super Mario montre que la façon dont le problème est présenté, et le but lointain qu'on vise ou non ont un profond impact sur la motivation. De même, l'idée de perdre quelque chose à chaque tentative infructueuse, aussi infime soit cette perte, a un impact non négligeable sur la motivation.

Donc, préciser et rappeler les buts ultimes (vous serez de bons scientifiques des données dans votre carrière) aide à surmonter les petites difficultés du quotidien. De même, nous devons éviter des cotations qui sanctionnent les étudiants par la perte de points, et promouvoir l'émulation par le gain de points à chaque fois que l'étudiant progresse, au contraire.

1.3.9 Teaching tech together

Dans son ouvrage, Greg Wilson pose un très grand nombre de principes qui fonctionnent bien pour l'enseignement de matières techniques. Au tout début, il définit dix règles :

1. Be kind: all else is details.
2. Remember that you are not your learners...
3. ...that most people would rather fail than change...
4. ...and that ninety percent of magic consists of knowing one extra thing.
5. Never teach alone.
6. Never hesitate to sacrifice truth for clarity.
7. Make every mistake a lesson.
8. Remember that no lesson survives first contact with learners...
9. ...that every lesson is too short for the teacher and too long for the learner...
10. ...and that nobody will be more excited about the lesson than you are.

Il y a énormément plus de choses intéressantes dans cet ouvrage, et il faudra prendre le temps de les étudier et digérer plus à fond plus tard !

1.3.10 Mort par PowerPoint

La présentation de type PowerPoint est partout et les enseignants ne peuvent plus s'en passer. Cependant, c'est un outil à double tranchant et une mauvaise utilisation de cet outil peut être désastreuse. C'est ce qu'on appelle la "mort par PowerPoint".

Quelques conseils pour améliorer sa présentation PowerPoint ici :

Don McMillan est un humoriste qui réussit à faire passer le message d'une manière hil-

rante ici :

1.3.11 Enseigner les statistiques comme les mathématiques

Douglas Andrews dans *the big mistake: teaching stat as though it were math* met en évidence le fait que les cours d'introduction aux statistiques qui sont trop orientés vers le côté mathématique éloignent les étudiants de l'essence même des statistiques et les font fuir. Au contraire, le côté appliqué des statistiques doit plutôt être mis en avant dans les cours introductifs.

La vidéo suivante caricature les interactions entre un biologiste et un statisticien. Ceci montre l'importance de former des personnes capables de faire le lien entre les deux disciplines : les scientifiques des données ayant un “background” en biologie.

1.3.12 Ludification de l'enseignement

L'apprentissage par le jeu est une approche intéressante et motivante. Le système de **badges** que l'on gagne en progressant offre des récompenses pour l'effort fourni. Il existe d'autres façons de rendre l'apprentissage plus ludique... et en la matière il faut essayer d'être imaginatif. Concernant la science des données, voici quelques pistes :

- **challenges** : les étudiants doivent répondre à un problème par équipe et soumettent des solutions à ce problème. La qualité des solution peut être quantifiée (ex : en machine learning, diverses métriques sont utilisables pour évaluer un classifieur). Un ranking est établit et affiché en temps réel. Cet exercice est très motivant et permet également de vérifier les acquis dans un exercice sommatif général.
- **course de relais** : un jeu de données à analyser, équipes de 4 (1h) ou 6 (1h30). Chaque étudiant a 1/4h pour travailler sur un dépôt GitHub puis fait un commit et passe le dépôt au suivant. Au bout du temps écoulé, le travail est évalué.
- **ping pong** : par exemple concernant la réalisation de graphiques. Par deux (ou deux binômes) qui s'opposent. On commence avec un chunk vide dans un document R Markdown ou Quarto pour poser le problème (avec un jeu de données à disposition). Le premier doit ajouter une instruction, puis le second, et ainsi de suite. Celui qui perd a soit ajouté une mauvaise instruction, soit l'analyse est terminée et il n'y a plus rien à rajouter (ou le délais max est écoulé et l'étudiant n'a rien pu rajouter).
- **correction contre la montre** : deux équipes s'affrontent. Chacune encode du code correct dans une app Shiny et propose une version avec une erreur

cachée à l'autre équipe. Les corrections sont soumises à quelque chose comme `gradethis::grade_code()` ou des outils de testing qui comparent deux versions de code, la réponse soumise par rapport à la version correcte. La première équipe qui trouve l'erreur a gagné.

- **quoi d'autre ? ...**

1.3.13 Let them eat the cake

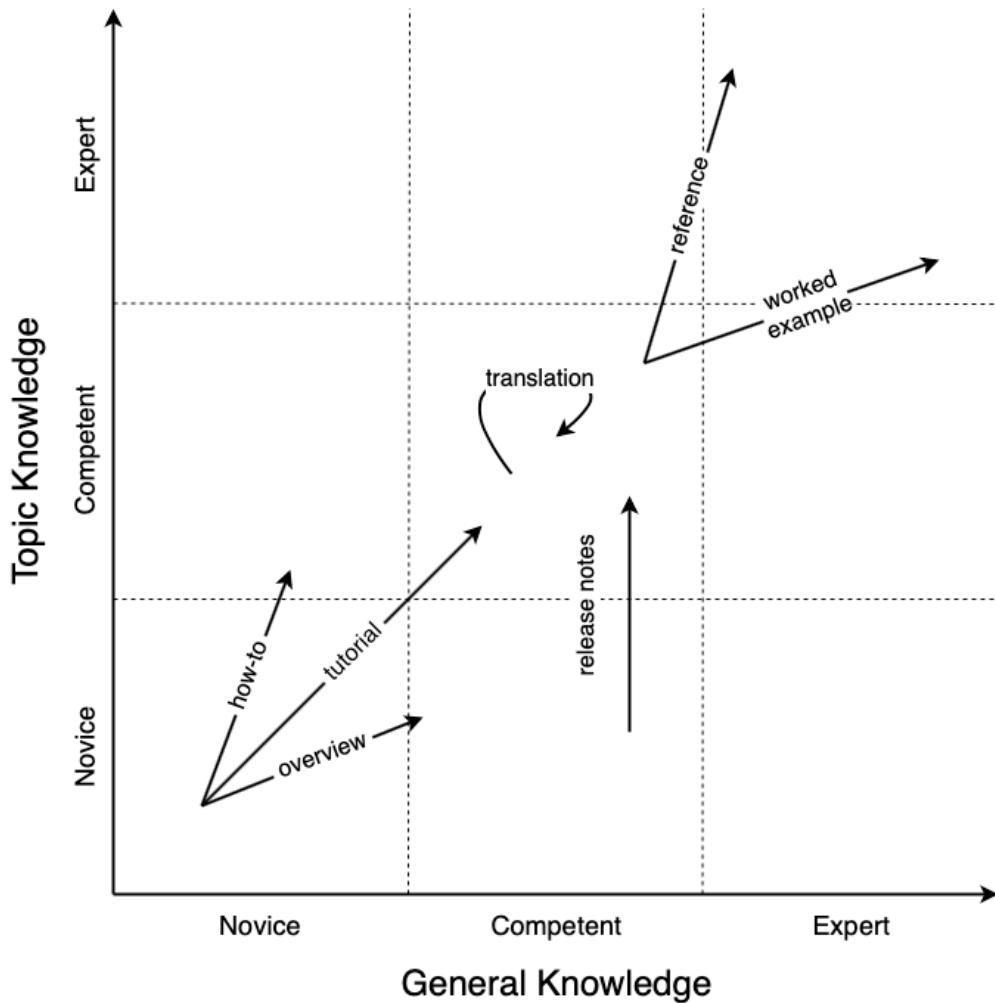
Le tout début d'un cours (la première heure, voire même les 10 premières minutes) sont déterminants pour la suite. Donc, il faut leur faire voir le résultat final et en tous cas, quelque chose de très motivant pour la suite. C'est la méthode prônée par Mine Cetinkaya-Rundel ici.

1.3.14 Differents type de documents

Dans documentation types Gregory Wilson propose une classification des différents types de documents en fonction du public et du but visé par le document.

- Le *novice* n'a pas encore de modèle mental du domaine. Il ne sait pas ce qu'il ignore et ne connaît pas le jargon.
- L'*utilisateur compétent* possède un modèle mental du domaine. Il peut exécuter des tâches routinières, peut trouver l'information et est capable de déterminer si ce qu'on lui présente est la solution à son problème ou non.
- L'*expert* a une vision dense et richement connectée du domaine. Il est capables de résoudre de nombreux problèmes facilement.

Le graphique suivant, tiré du blog de Gregory Wilson présente différents types de documentations en fonction des connaissances générales (axe X) *versus* les connaissances spécifiques sur la question posée (axe Y). Les flèches indiquent dans quelle direction la documentation amène cet utilisateur. Les types de documentations repris dans le schéma sont décrits plus en détails dans le blog d'origine.



Il faut noter que des documents de type **livre** ou **vignette** ne sont pas inclus dans le diagramme car en réalité, ce sont des documents composites qui contiennent généralement plusieurs sections apparentées à l'un des types repris dans le diagramme. L'**aide-mémoire** est un type particulier de documentation de référence synthétisée en une ou deux pages. Enfin les **bloc-notes informatiques** (comme le R Notebook) sont caractérisés par l'inclusion de code exécutable directement dans le document, mais à part cela, ils peuvent être de n'importe quel autre type.

Par rapport à l'enseignement de la science des données, ce sont les trois types qui visent le novice qui nous intéressent en premier :

- Le **tutoriel** correspond au type de document à privilégier dans notre matériel pédagogique. Son objectif est de **construire un modèle mental du domaine** ainsi qu'à acquérir quelques compétences de bases permettant de progresser et devenir un utilisateur compétent.
- L'**overview** (vue d'ensemble) peut être un webinaire, un premier cours, un “elevator pitch”, du matériel de marketing... Il informe essentiellement de l'existence de quelque chose, mais sans apporter de compétences supplémentaires (l'utilisation n'apprend pas à résoudre des problèmes). Par rapport au schéma d'apprentissage,

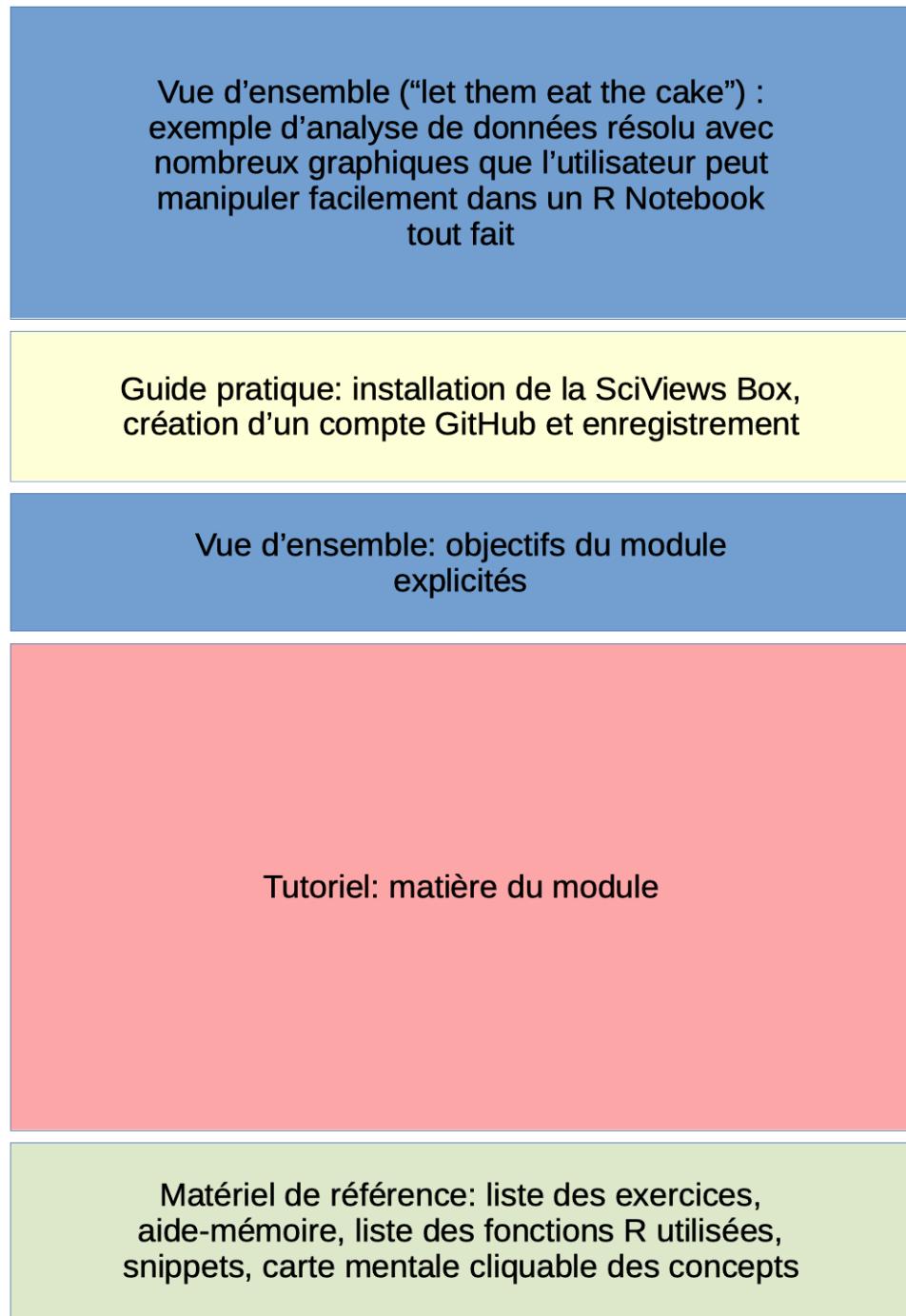
ce type de document est utile (et même indispensable) pour éveiller l'intérêt et donc, pour capter l'attention de l'étudiant.

- Le **how-to** (guide pratique) est une recette étape par étape pour résoudre un problème particulier. Il n'est pas nécessaire de comprendre le pourquoi pour pouvoir l'utiliser. Par contre, c'est le type de documentation qui permettra de résoudre un problème particulier le plus vite possible, mais ne change rien aux connaissances générales du domaine de l'utilisateur.

Nous voyons donc que les vues d'ensemble et les guides pratiques sont à réserver à des situations particulières : éveil de l'intérêt en début de cours ou de module pour les premiers, résoudre rapidement un problème pour pouvoir passer à autre chose pour les seconds (par exemple : installer le logiciel requis le plus vite possible pour pouvoir se concentrer sur les exercices ensuite). Sinon globalement, le matériel pédagogique doit se structurer essentiellement comme un tutoriel.

Cependant, il ne faut pas négliger non plus de la documentation complémentaire qui sera utile lorsque l'étudiant deviendra compétent : le matériel de **référence** (dont les aides-mémoires, foires aux questions, cartes mentales, aides en ligne des fonctions de R) et les résolutions de cas (**worked examples** dont les check-listes et les snippets font partie). En effet, une fois qu'il a assimilé les notions, le document de type tutoriel qu'il a utilisé jusqu'ici ne présente *pas* l'information d'une manière telle qu'elle soit facile à retrouver. Les points importants sont trop délayés (développement étape par étape, mélange d'exemples et de théorie, etc.). Donc, il faut *résumer* les notions importantes dans des documents de référence également fournis aux étudiants et leur apprendre à les utiliser et à passer des uns (tutoriels) aux autres (références). **Fournir seulement un syllabus bien construit comme un tutoriel ne suffit pas car le novice devenu utilisateur compétent à son usage se retrouve alors face à un document qui est inadapté comme documentation de référence.** Une approche à expérimenter serait de faire élaborer (au moins partiellement) le matériel de référence par les étudiants. Ceci va dans le sens de la pédagogie active et du concept de classe renversée, mais tout en guidant plus fortement la matière à étudier puisqu'on part d'un syllabus/tutoriel qui cadre le travail.

La scénarisation du cours est importante, mais également la scénarisation du matériel pédagogique. Un point de vue qui paraît essentiel ici est de bien définir le type de documentation pour chaque partie d'un chapitre *avant* de l'écrire et ensuite de se tenir au style et aux objectifs relatifs à ce type de documentation. L'illustration suivante est une tentative de scénarisation d'un module en fonction de quatre **types de documents**, présentés ici en couleurs différentes. Ceci semble complémentaire à la définition du contenu en terme de **notions à enseigner** et d'**acquis d'apprentissage** visés par le module qui définissent *quoi* enseigner, alors que la présentation par types de documents indique *comment* l'enseigner.



1.4 Organisation de la matière

Le découpage de la matière dans les cours de science des données, ainsi que le timing sont importants pour obtenir un **travail sur la durée des étudiants tout au long de**

l'année qui est une autre règle majeure pour un apprentissage efficace et durable. Voici quelques idées à développer pour assurer cet apprentissage progressif et continu.

1.4.1 Découpage

- **Module** Nous avons entre 6 et 12 modules selon les cours. Chaque module demande entre 10 et 15h de travail dont 6h en présentiel. Il s'étale sur une semaine (minimum). Un timing classique est : 3h de préparation à domicile, présentiel de 2 ou 4h, travail à nouveau à domicile 1h, présentiel de 4 ou 2h, et enfin, travail à domicile 2h pour finaliser le tout. Au niveau de chaque module, on précise les objectifs, et en fin de module, il faut un bilan avec auto-évaluation pour que l'étudiant puisse vérifier qu'il ait bien compris les concepts abordés de lui-même (ex. : summary de H5P)
- **Capsule** Un ensemble plus restreint d'items à apprendre (généralement un ou deux items max). Section du bookdown et/ou vidéo avec des exercices H5P/learnr niveau I entrelacés. Préciser également l'objectif et faire un mini-bilan à la fin de la capsule.
- **Tâche** Item d'apprentissage unique lié à un et un seul matériel pédagogique (sous-section bookdown, présentation H5P, mini-learnr, très courte vidéo ou gif animé).
- **Activité** Exercice réalisé par l'étudiant. Inclus dans les items précédents.

1.4.2 Quatre niveaux de difficulté

Au fil de l'élaboration de notre matériel pédagogique diversifié, nous avons observé que certains formats sont plus adaptés pour des niveaux de difficulté plus simples, et d'autres pour des tâches plus complexes. Par exemple, des petits widgets H5P simples sont utiles pour garder l'étudiant actif au milieu de la lecture d'un texte un peu long, autant qu'ils permettent à l'étudiant d'auto-évaluer s'il a bien compris. Par contre, les jeux de données à analyser soi-même avec des directives minimalistes dans des projets GitHub en groupe sont d'un niveau de difficulté nettement plus élevé. Au final, nous définissons quatre niveaux de difficulté croissants, et nous leurs associons des matériaux et des objectifs différents. **L'idéal est de balayer un même concept au travers de matériel appartenant chaque fois à l'un de ces quatre niveaux de difficulté pour assurer la gradation de difficulté et la répétition non monotone des concepts, facteurs clés d'un apprentissage actif réussi.**

1.4.2.1 Niveau 1

Exercices simples directement inclus dans le bookdown.

- Objectif : rendre l'étudiant actif et lui permettre une auto-évaluation de la compréhension correcte des concepts en direct.
- Type : H5P, learnr avec 1 ou 2 exercices, applications Shiny simples directement intégrés dans le bookdown.
- Code R : pas d'écriture de code directement, éventuellement utiliser des widgets H5P du genre remplir les blancs dans du code à partir d'une liste de fragments, indiquer si un code présenté est correct ou non, s'il effectue un traitement donné ou pas (question true/false), ...
- Lieu : travail à domicile.
- Suivi de l'utilisation des outils dans le rapport de progression.

1.4.2.2 Niveau 2

Auto-évaluation des principes théoriques à semi-pratiques par l'étudiant via des learnrs comportant plusieurs questions.

- Objectif : pouvoir vérifier ses acquis, faire un bilan de progression en auto-évaluation.
- Type : learnrs exécutés soit dans le bookdown, soit dans RStudio, à voir...
- Code R : écriture minimale de code via des exercices guidés.
- Lieu : travail à domicile et/ou en présentiel.
- Suivi avec acquisition de badges. Les évaluations plus “théoriques” passent aussi par des learnrs. Question : évaluation en fin d'AA pendant la période d'examen, ou évaluation en continu ?

1.4.2.3 Niveau 3

Projets GitHub individuels cadrés.

- Objectif : effectuer une transition douce vers l'application pratique des concepts. Analyse de données réelles fortement guidée.
- Type : projets RStudio avec fichiers partiellement remplis, tâches bien précises et nombreux commentaires d'aide dans les fichiers “template” fournis.
- Code R : écriture de code par rapport à un résultat attendu bien cadre et spécifié. Essentiellement carnet de notes, mais d'autres formes (rapport, présentation, article...) peuvent également être abordées brièvement à ce stade.
- Utilisation d'une batterie de tests pour que l'étudiant puisse avoir un retour immédiat sur ce qu'il a réalisé correctement ou non, et dans ce derniers, avoir des conseils pour corriger ses erreurs.
- Évaluation par grille critériée remplie de manière semi-automatique en s'appuyant sur les mêmes tests que ci-dessus.
- Lieu : travail en présentiel essentiellement en individuel.
- Évaluation et retour de temps en temps via l'inspection des dépôts GitHub des étudiants. Historique du travail de chaque étudiant via Git. Retour important ici (s'astreindre à visualiser quelques dépôts chaque semaine et faire des retours via issues GitHub, mails, ou Discord).

1.4.2.4 Niveau 4

Analyse libre de données dans des projets GitHub en groupe.

- Objectif : acquérir un certain degré d'autonomie dans l'analyse des données en pratique.
- Type : projet RStudio avec instructions réduites au minimum, pas de fichiers partiellement préremplis.
- Code R : écriture de code en fonction des objectifs fixés par l'étudiant lui-même. Il est aussi responsable de l'organisation et de la maintenance de ce code jusqu'à l'aboutissement du projet. Présentation sous forme de carnet de notes, et ensuite sélection des éléments les plus pertinents pour rédiger un rapport qui doit être “présentable” à un panel de lecteurs.

- Travail par groupe de deux ou quatre étudiants pour mobiliser l'*intelligence collective* et permettre aux étudiants plus faibles de progresser grâce aux apports des plus forts dans le groupe.
- Lieu : travail en présentiel à compléter à domicile.
- Les rapports font l'objet d'une évaluation détaillée avec grille critériée et sont la base de la note pour la partie pratique. Historique de l'activité via Git. Typiquement, un gros projet qui reprend les notions de plusieurs modules sur un quadrimestre.

1.4.3 Timing

Les différents modules sont placés à l'horaire à des semaines bien définies. L'étudiant doit prendre conscience qu'il a une semaine pour assimiler le contenu du module, tout en lui permettant une certaine flexibilité (aussi pour les étudiants en passerelle qui arrivent plus tard et qui doivent pouvoir rattraper *progressivement*). Éventuellement, pour eux il faudrait pouvoir établir un calendrier adapté.

1.4.4 Suivi

Voici les éléments que nous souhaitons mettre en place pour assurer un apprentissage dans la durée, et le décrochement d'un minimum d'étudiants.

- Calendrier d'achèvement souhaité des modules
- Rapport de progression le plus en temps réel possible. Comparaison à la fois avec le calendrier et avec la progression générale de la classe (motivation).
- Acquisition de badges une fois les modules achevés, après mini-examen de type learnr (motivation).
- Visualiser le calendrier souhaité sur le rapport de progression. Également, visualiser l'avancement global de l'étudiant par rapport à ce qui est attendu.
- Le rapport de progression se bloque si un module est “by-passé” par l'étudiant : sa progression dans les modules ultérieurs n'est pas affichée tant qu'il n'a pas terminé le module “by-passé”. Cela force à suivre le parcours de progression module par module. Ceci est surtout important pour SDD I, où les étudiants sont confrontés à une approche pédagogique à laquelle ils ne sont pas habitués, mais à voir pour les autres cours également.
- Obtention des points en fonction de la progression. Aussi, une évaluation du rapport d'analyse niveau IV pour la note de l'AA qui complète la note que l'étudiant obtient. Nous n'aurons donc plus d'examen pendant la session qui ne fait que traduire la capacité de “bachotage” de la matière par l'étudiant, alors que notre objectif est un apprentissage continu, durable et par la pratique autant que possible.
- Possibilité d'obtention de badges “spéciaux”, liés à des points bonus : pousser et motiver les étudiants dans différentes directions (entre-aide, excellence, esprit d'initiative, ...).

- Émulation par le jeu. Varier les versions. Par exemple, la première équipe qui trouve la réponse a gagné. Aussi les compétitions de type “challenge kaggle” pourraient être réalisées pour la partie modélisation et la partie machine learning. Un jeu de test auxiliaire secret est utilisé par les enseignants pour quantifier la qualité du modèle proposé par les étudiants et permet de réaliser un classement non subjectifs des équipes en compétition. Les étudiants soumettent autant de modèles qu’ils le souhaitent et un classement des équipes est réalisé pour le fun (ou alors, récompenses à la clé sous forme d’un point bonus, de chocotofs, voire de choses plus sérieuses comme une clé USB par exemple).

Chapter 2

Organisation générale

La plateforme pédagogique LearnIt::R est constituée d'un ensemble d'outils logiciels et de services permettant de créer des cours interactifs en ligne. Dans sa version complète, elle est complexe. Heureusement, vous pouvez vous approprier ses différentes fonctionnalités progressivement. Une version centrée sur GitHub, appelée pour cette raison **LearnIt::R GitHub**, ne nécessite pas l'installation de serveurs spécialisés, ni de base de données. Nous conseillons de commencer par cette version. **LearnIt::R LRS** est une version plus complète qui nécessite l'installation d'une base de données (LRS pour *Learning Record Store*) pour collecter les traces de l'activité des étudiants en temps réel. Cette base de données ouvre d'autres possibilités qui sont exploitées par cette mouture plus complète de la plateforme. Par la suite, ces deux versions sont décrites successivement.

Avant d'installer la plateforme LearnIt::R, vous devez configurer un environnement logiciel basé autour de R et de RStudio pour vous, mais aussi pour vos étudiants (si l'apprentissage de R fait partie de votre cours).

- R est un logiciel Open Source de statistique, visualisation et manipulation des données. C'est le langage principal des outils composant LearnIt::R. Si votre cours vise, entre autres à l'enseigner R à vos étudiants, vous n'aurez pas à apprendre un autre langage pour accéder aux fonctionnalités de la plateforme. R est un logiciel puissant et gratuit que l'étudiant pourra s'approprier pour faire ensuite ses analyses tout au long de sa carrière s'il le souhaite. Il est détaillé au chapitre 3.
- RStudio est un environnement de développement intégré (IDE pour “*integrated development environment*” en anglais) gratuit. Il permet aux étudiants de réaliser leurs exercices en R facilement. Une version Desktop est installable localement sur un ordinateur. La version serveur permet de travailler dans une interface utilisateur identique à peu de choses près sur le cloud. Couplé à R Markdown ou Quarto, RStudio offre la possibilité d'édition du code R et d'écriture de rapports ou autres documents dans un environnement unique. Son installation et utilisation sont discutés au chapitre 3.

2.1 LearnIt::R GitHub

La plateforme LearnIt::R GitHub se base essentiellement sur des solutions Open Source et sur GitHub. Elle utilise R et RStudio décrits plus haut, mais aussi GitHub et GitHub

Classroom. Les packages R `{quarto}`, `{learnnr}`, `{shiny}` et `{learnitdown}` forment la base des fonctionnalités de LearnIt::R. Pas de panique si tous ces termes vous sont inconnus : nous allons les expliquer l'un après l'autre. Leur utilisation sera ensuite détaillée chapitre après chapitre dans la section consacrée à LearnIt::R GitHub pour que vous puissiez implémenter la plateforme de manière graduelle en maîtrisant un outil après l'autre.

- Git est un système de gestion de version largement utilisé en programmation. Combiné à R, RStudio et R Markdown, il permet de travailler de manière *reproductible*. L'utilisation de git peut être assez complexe, mais en se restreignant à des actions simples et aux outils git disponibles dans RStudio, il est possible de l'utiliser dans vos cours. La maîtrise, même minimale, de git est un plus pour les étudiants qui apprennent et utilisent le langage R. L'objectif est d'amener vos étudiants à prendre de bonnes habitudes en matière de traçabilité et reproductibilité de leurs analyses. GitHub est un hébergement et lieu d'échange de dépôts git sur le cloud. Une bonne partie des ses fonctionnalités sont accessibles gratuitement. Il est puissant, universellement utilisé et il est couplé à RStudio. L'interface git et GitHub de RStudio rend leur utilisation plus simple. GitHub Desktop pourra éventuellement compléter l'ensemble pour effectuer des tâches non reprises dans l'interface git de RStudio s'il n'est pas prévu de recourir à la ligne de commande (qui est tout de même la manière la plus efficace d'utiliser git). Enfin, GitHub Classroom est une extension permettant de gérer les dépôts créés pour vos étudiants dans le cadre de vos cours. Ces différents outils sont présentés au chptire 4.
- Bookdown est un type de sortie R Markdown multi-documents qui permet d'intégrer également des calculs et graphiques de R. Vous pouvez l'utiliser comme moyen de centraliser facilement tout votre matériel pédagogique et l'organiser en modules. Ces modules formeront autant de chapitres différents du bookdown. LearnIt::R propose, par l'intermédiaire du package `{learnitdown}`, des fonctionnalités supplémentaires pour y intégrer vos exercices ou du contenu contextuel (voir 5). et notamment des applications Shiny (voir 7) et des widgets H5P (voir 8). Le chapitre 5 y est donc consacré.
- Shiny permet de créer des applications sous R avec interface graphique dans un browser web. Les possibilités sont pratiquement sans limites pour créer des applications interactives et démontrer les concepts "en live" à vos étudiants. Les applications Shiny sont intégrables dans bookdown, mais elles doivent être exécutées soit sur un serveur dédié, soit localement sur l'ordinateur de l'étudiant. Un type particulier d'application Shiny est constitué par les learnrs. Ces documents interactifs de type tutoriels présentent des exercices dans R et des choix multiples aux étudiants. Ces tutoriels learnrs forment la base des outils d'apprentissage à domicile dans LearnIt::R. La correction automatique des exercices learnr est possible grâce à `{gradethis}`. Le chapitre 6 détaille l'utilisation des tutoriels tandis que le chapitre 7 est consacré à Shiny.
- H5P propose une large variété de contenus javascript intégrables dans une page web et donc dans vos contenus bookdown pour varier et rendre plus efficace l'apprentissage. Sans l'installation d'un serveur dédié, vous pouvez créer du contenu sur H5P.org ou utiliser un service payant plus complet sur H5P.com. Si votre institution utilise Moodle, H5P fournit un plugin qui permet d'intégrer ce type d'exercice dans Moodle, mais alors, vous ne pourrez pas les placer directement dans votre bookdown (vous utiliserez un lien qui ouvrira la ressource correspondante dans Moodle). Voyez le chapitre 8 pour plus de détails.

- Enfin, le package `{learnitgrid}` a été écrit spécialement pour faciliter la correction des rapports de vos étudiants dans les projets GitHub Classroom. Le chapitre 9 est entièrement consacré à l'utilisation de GitHub Classroom et de `{learnitgrid}` pour gérer les projets des étudiants.

2.2 LearnIt::R LRS

La version complète de la plateforme LearnIt::R, appelée LearnIt::R LRS, nécessite l'installation d'une base de données MongoDB pour collecter les traces de l'activité des étudiants en temps réel. Cette base de données ouvre d'autres possibilités pour collecter, visualiser et traiter les informations relatives à la progression de vos étudiants. Naturellement, des adaptations des autres outils sont nécessaires pour les faire fonctionner avec cette base de données. Vous pourrez utiliser MongoDB sur vos propres serveurs, ou utiliser MongoDB Atlas, voire même, les deux ensemble. Tout cela est expliqué au chapitre 10.

- Le matériel du cours en ligne écrit en bookdown doit être intégré dans un serveur Wordpress qui implémente des fonctionnalités côté serveur pour l'identification des utilisateurs et l'envoi des données à la base de données MongoDB. Vos contenus H5P sont alors gérables depuis Wordpress et un plugin H5PxAPIkatchu offre la possibilité d'enregistrer l'activité de vos étudiants dans les exercices H5P intégrés au bookdown dans la base de données MongoDB. Voyez le chapitre 11 pour son installation et utilisation.
- Moodle est une plateforme d'apprentissage en ligne largement utilisée dans diverses institutions. LearnIt::R LRS peut récupérer (de manière optionnelle) les informations relatives à vos étudiants à partir de Moodle pour compléter sa table des utilisateurs. L'étudiant doit lier son cours à Moodle avant de faire quoi que ce soit d'autre, comme expliqué au chapitre 12.
- Un serveur Posit Connect peut être installé (payant et assez cher, mais la version enseignement est gratuite) pour servir vos différents applications Shiny et learnr. Le chapitre 13 discute de son utilité dans le cadre de la plateforme LearnIt::R LRS.
- Un rapport de progression à l'usage des étudiants est implémenté dans le package `{learnitprogress}`. Il est utile pour voir où l'on en est dans son apprentissage. Voyez le chapitre 14 pour son utilisation.
- Enfin, un tableau de bord destiné aux enseignants est fourni par le package `{leanitdashboard}`. Il permet le suivi de l'apprentissage du côté enseignant et il est utile également pour repérer rapidement les étudiants en décrochage ou les autres problèmes. Le chapitre 15 est consacré à son utilisation.



Ces nombreux outils et services devant collaborer entre eux pour fournir un environnement d'apprentissage complet et cohérent, il est important de les configurer et de les inter-connecter correctement. La façon de le faire est détaillée dans la seconde partie de cet ouvrage consacrée à LearnIt::R LRS.

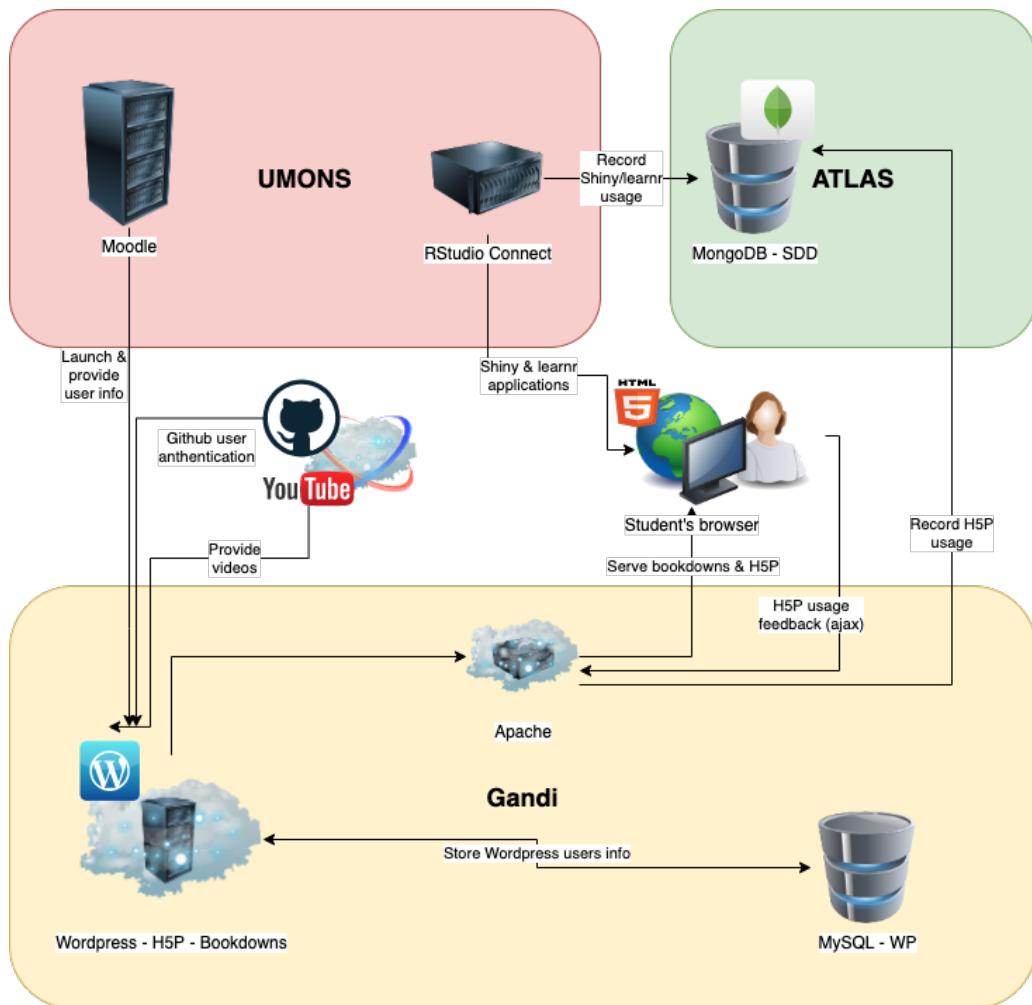


Figure 2.1: Vue générale des logiciels utilisés pour implémenter LearnIt::R LRS.

2.3 Outils supplémentaires

- Les vidéos relatives à vos cours pourront être hébergées sur Youtube, voyez par exemple le canal BioDataScience_Course. Le chapitre 16 explique en détails comment créer ces vidéos et avec quels outils.
- Les exercices de type QCM ou autres peuvent être créés avec Wooclap. Ce service en ligne permet de créer des exercices interactifs que vous pourrez intégrer dans vos cours. Les résultats des étudiants sont collectés et peuvent être visualisés en temps réel. Le chapitre 17 est consacré à son utilisation.

2.4 RGPD et droit d'auteur

Les questions relatives aux données d'un point de vue plus légal sont traitées ici.

2.4.1 RGPD

Le **Règlement Général sur la Protection des Données** ou RGPD (n° 2016/679) adopté par l'Union Européenne en 2016 régit la façon dont les données à caractère personnel peuvent être collectées et utilisées. Ce règlement est en faveur de l'utilisateur (ici l'étudiant). Il faut notamment son accord pour collecter et utiliser ses données personnelles. Or de telles données sont indispensables pour suivre la progression des étudiants, pouvoir les noter, etc.

Par exemple à l'UMONS, l'étudiant signe le document adéquat lors de son inscription (charte de la vie privée, complétée de documents disponibles sur l'Intranet UMONS). Ce document mets **Moodle** en conformité par rapport au RGPD, mais pas les outils externes, tels que ceux de LearnIt::R. Nous devons donc préciser exactement ce que nous devons faire pour être en conformité RGPD par l'intermédiaire d'un document complémentaire affiché sur le site (entrée Protection de la vie privée dans le menu à gauche) qui informe et précise la façon dans les données à caractère personnel des étudiants sont utilisées. Et puisque ces données entrent dans un cadre d'utilité publique et que l'usage est similaire à ce qui se fait sur Moodle, un accord supplémentaire (au delà des documents qu'il a déjà rempli à son inscription) n'est pas nécessaire de la part de l'étudiant.

L'utilisateur doit avoir la possibilité d'effacer intégralement ses données personnelles s'il le souhaite lorsqu'il efface son compte d'un site. C'est indiqué explicitement dans le RGPD. Cependant, des restrictions à ceci sont indispensables pour la bonne gestion du suivi des étudiants à l'Université. C'est pour cela que l'étudiant a du signer un accord lors de son inscription. Afin d'être en conformité avec cette directive, nous avons rajouté un bouton **Effacer mes données personnelles** dans la première page des cours bookdown. S'il clique sur ce bouton, l'utilisateur peut ensuite lire le contenu de manière anonyme et les activités learnr/Shiny ne sont **pas** enregistrées. Par contre, H5P est toujours enregistré et les données historiques antérieures sont toujours dans la base de données.

L'anonymisation ou la pseudonymisation des données est de mise lorsque ces données servent à une étude générale (par exemple, évolution des performances des cohortes d'étudiants avec les outils progressivement mis en place, études scientifiques, ...). Cela passe par l'effacement des données de toute information à caractère personnel. Le nom, numéro de matricule ou adresse email de l'étudiant sont remplacés par un identifiant générique, par exemple, "étudiant 1", "étudiant 2", etc. Il faut mettre cette pratique

en œuvre pour toutes les études rétrospectives générales visant à estimer l'impact pédagogique des outils mis en place (voir appendice C).

Lien utiles

- RGPD pour les développeurs par la CNIL en France,
- Techniques d'anonymisation
- Recommandation relative aux mots de passe
- Présentation de l'information RGPD aux utilisateurs

2.4.2 Droit d'auteur et licence

TODO: discuter le choix de la licence pour le contenu en ligne.

TODO: mise en œuvre d'un outil anti-plagiat.

TODO: aspects relatif à l'IA et aux LLM.

Chapter 3

R & RStudio

TODO: expliquer les différentes façons d'avoir un système opérationnel avec R + RStudio + packages requis :

- Installation en local, Windows, macOS et Linux
- Utilisation de RStudio Cloud
- SaturnCloud
- VirtualBox
- Docker/Podman
- SciViews Box
- Avantage et inconvénients des différentes solutions

(texte à modifier relatif à VirtualBox et la SciView Box)

- La SciViews Box est une machine virtuelle sous VirtualBox totalement préconfigurée. Nous créons une SciViews Box chaque année avec tous les logiciels nécessaires aux étudiants mais pas seulement : il s'agit aussi d'une machine virtuelle permettant de réaliser de nombreuses analyses de façon parfaitement reproductible. Faire installer tous les logiciels nécessaires par les étudiants n'est pas possible. Plusieurs solutions s'offrent à nous pour proposer un environnement logiciel préconfiguré pour les cours :
 - Utiliser RStudio cloud serait la solution la plus simple au départ car l'étudiant n'a qu'à se connecter et c'est parti. Même avec un Chromebook, il peut travailler sur sa session RStudio Cloud. Cependant, la session n'est disponible *que* pour le cours et l'étudiant doit tout installer sur son ordinateur ensuite s'il veut continuer à utiliser R, RStudio et Git/GitHub (par exemple, dans le cadre de son mémoire). Or précisément, il n'a jamais appris à installer ces outils. Cela nous paraît être une barrière importante à l'adoption de ces outils logiciels au-delà du cours, même si beaucoup de (grandes) universités ont choisi cette voie.

- Utiliser Docker et Rocker qui est une solution de containerisation pour R et RStudio. Si ces outils fonctionnent bien en production, nous les avons trouvés moins pratiques que la machine virtuelle pour nos étudiants, surtout dans un contexte où la quasi-totalité des ordinateurs utilisés sont sous Windows ou Mac OS (Docker fonctionne particulièrement bien dans un environnement Linux au départ). L'installation et la gestion de containers Dockers est simple pour un utilisateur pointu, mais nous, nous démarrons avec certains étudiants qui ont une connaissance bien moins technique de l'informatique. La moindre action d'installation ou de configuration, aussi simple soit-elle, paraît démesurément complexe à leurs yeux. Il faut donc démarrer avec de l'**ultra-simple**, autrement dit, fonctionnel directement ou s'auto-installant, auto-configurant tout seul.
- Déléguer le calcul sur le cloud à travers Binder. Aujourd'hui, il est très facile d'exécuter du code dans un document Jupyter ou R Markdown via ce service gratuit sur le cloud. Et c'est effectivement d'un accès ultra-simple du côté de l'utilisateur qui n'a qu'à cliquer sur un lien pour lancer l'application. Au moins un format de cours en ligne y recourt (template par Ines Montani, voir par exemple pour un cours concret en R, Supervised Machine Learning Case Studies in R). Nous l'avons testé, mais l'avons jugé à l'usage trop lent au démarrage et trop limité ensuite. De plus, comme pour RStudio Cloud, les étudiants n'ont aucun logiciel utilisable sur leur PC pour travailler ensuite. La simplicité au départ ne fait donc que repousser la difficulté ultérieurement de devoir quand même installer R, RStudio, git, etc. sur son PC.
- Enfin, la solution de la machine virtuelle avec la SciViews Box nous permet d'offrir une série de logiciels préconfigurés en version serveur et/ou locale avec interface graphique sous Xubuntu. C'est une bonne base, mais l'installation était encore un peu laborieuse lorsqu'elle devait se faire à la main dans VirtualBox (on en revient pratiquement à des difficultés similaires à Docker, par exemple). Nous avons donc conçu des logiciels annexes d'installation et de lancement de la machine virtuelle de sorte que celle-ci ne soit qu'à un clic de souris (ou presque) pour nos étudiants débutants ! A l'usage, cela fonctionne très bien. Nous avons un lanceur rapide par édition de SciViews Box, et ce, à la fois sous Windows et sous Mac OS. L'étudiant a un logiciel réellement installé sur son PC, qu'il peut parfaitement continuer à utiliser au delà des cours eux-mêmes.

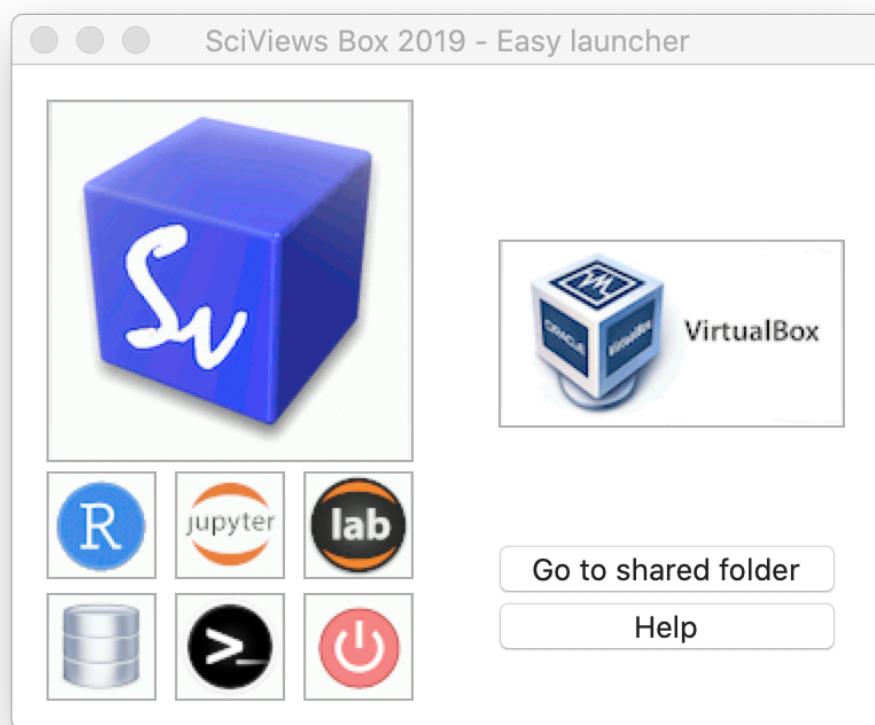


Figure 3.1: Lanceur rapide de la SciViews Box 2019

Part I

LearnIt::R GitHub

Chapter 4

Git & GitHub

Nous utilisons abondamment et enseignons l'usage de git et GitHub dans le cadre de nos cours de **Science des Données Biologiques** à l'UMONS. De plus, nous utilisons GitHub Classroom pour gérer la distribution des travaux aux étudiants sous forme de dépôts GitHub. Cette section couvre ces logiciels.

4.1 Installation Git et cie

Dans la SciViews Box, git, GitHub Desktop et RStudio server (incluant une interface vers git) sont préinstallés. Hors machine virtuelle, il faut installer git, RStudio et GitHub Desktop. Le reste se fait en interagissant avec GitHub via le navigateur web.

4.2 BioDataScience-Course

L'organisation GitHub BioDataScience-Course regroupe tous les dépôts relatifs à nos cours, y compris les travaux des étudiants.

4.2.1 Convention de noms

Comme l'organisation contient plusieurs milliers de dépôts, il est utile de les nommer correctement pour s'y retrouver. Les dépôts créés via GitHub Classroom seront traités plus loin. Voici quelques règles :

- Les dépôts important (matériel pédagogique, gestion des étudiants et de leur progression, ...) commencent tous par `sdd` lorsque ces dépôts sont utilisés en interne (ex. : `sdd_lesson`, `sdd_movies`, `sdd-admin`, ...). **Attention : `sdd` est suivi d'un trait souligné (`sdd_`) pour les dépôts classiques comme par exemple `sdd_lesson`, mais d'un tiret (`sdd-`) pour les bookdowns comme par exemple `sdd-umons-2020`. Cette distinction est importante pour éviter de mélanger ces deux types de dépôts.
- Les dépôt concernant des items publics peuvent aussi utiliser `BioDataScience` comme préfixe pour être plus explicite (ex. : `BioDataScience1`, `BioDataScienceData`, `BioDataScienceExam`, ...). Pour les dépôts moins importants ou de test, utiliser un autre préfixe.

- Lorsqu'un dépôt est relatif à une cohorte d'étudiants en particulier, nous l'indiquons par un suffixe reprenant les deux derniers chiffres de l'année initiale, suivi de `m` pour les étudiants montois et `c` pour les étudiants de Charleroi. Par exemple, `sdd_students19m` concerne la cohorte d'étudiants montois inscrits en premier cours (Bab2 de biologie) en 2019-2020. De même `sdd_students20c` concerne la cohorte d'étudiants de Charleroi inscrits en Bab2 en 2020-2021.

4.3 GitHub Classroom

Nous utilisons GitHub Classroom pour gérer la distribution des travaux aux étudiants sous forme de dépôts GitHub. GitHub Classroom se lie à une organisation GitHub (dans notre cas, `BioDataScience-course`). Il permet de gérer différentes classes afin d'organiser les cohortes d'étudiants. Les conventions de noms sont similaires aux conventions présentées précédemment. Par exemple, la classe `sdd1-19m` correspond au cours de Science des données biologiques I donné aux étudiants montois de l'année académique 2019-2020. **Commentaire PhG : ne devrait-on pas plutôt utiliser `sdd_19m` pour mieux respecter les conventions ci-dessus ? Mais problèmes avec le sous-tiret dans GitHub Classroom ?**

Au sein de chaque classe, il est possible de créer des exercices individuels ou par groupe que l'on nomme des assignations (*assignments* en anglais). Chaque nom d'assignation débute par le code lié à la classe. Cette nomenclature permet une gestion simplifiée de l'organisation sur GitHub. Par exemple, l'assignation `sdd2-20m-mod8` porte sur un exercice ciblé sur le module 8 du cours de science des données biologiques 2 pour les étudiants montois suivant ce cours durant l'année académique 2020-2021. Si le dépôt est complexe et porte sur plusieurs modules, il porte un nom lié au travail comme `sdd2-20m-biometry`.

Les exercices peuvent se baser sur un dépôt modèle (*template* en anglais) afin de guider les étudiants. Nous fournissons toujours un dépôt template pour nos exercices. Un dépôt template est un dépôt GitHub reconnu comme *Template repository* (ce paramètre est sélectionnable dans la configuration du dépôt).

Parmi les paramètres de l'exercice, il est important de spécifier si le dépôt est privé ou public. Par défaut, les dépôts reprenant les exercices des étudiants sont privés, mais chaque étudiant est libre de basculer n'importe lequel de ses dépôts en public afin d'offrir une visibilité à ce qu'il a réalisé dans R (peut-être important en matériel complémentaire à un C.V. pour des profils plus typés science de données).

Un lien URL est fourni pour chaque exercice afin d'inviter les étudiants à rejoindre l'exercice. Pour les exercices individuels, GitHub Classroom va créer un dépôt généré à partir du dépôt template dans l'organisation `BioDataScience-Course` portant le nom de l'exercice suivi du nom d'utilisateur GitHub de l'étudiant (par exemple, `sdd2-20m-mod8-GuyliannEngels`). Les travaux de groupe portent le nom de l'exercice suivi du nom du groupe.

GitHub Classroom est en évolution constante. De nouvelles fonctionnalités apparaissent régulièrement afin d'améliorer la gestion des classes et des exercices. Par exemple l'intégration de GitHub Classroom avec Moodle est en cours de développement. GitHub Classroom propose également l'application GitHub Classroom Assistant afin de télécharger l'ensemble des dépôts liés à un exercice dans un dossier de notre choix. Bien pratique pour les enseignants au moment d'évaluer les travaux des étudiants.



Une autre ressource intéressante dans le contexte de l'utilisation de GitHub/GitHub Classroom pour l'enseignement est GitHub Education.

4.4 Astuces Git et GitHub

Quelques astuces qui nous sont utiles...

- Comparaison de branches sur GitHub : ajouter `/compare` à la fin de l'URL du dépôt et une page de comparaison de branche apparaît. Sinon, cette comparaison est aussi possible dans GitHub Desktop (menu `Branch` → `Compare to Branch`).
- Lorsqu'une branche n'est plus utile, elle peut être facilement éliminée depuis GitHub Desktop (entrée de menu `Branch` → `Delete...`). Cette branche doit être activée d'abord, et tous les commits doivent avoir été pushés. Il est conseillé d'éliminer la branche sur le serveur en même temps que la branche locale.

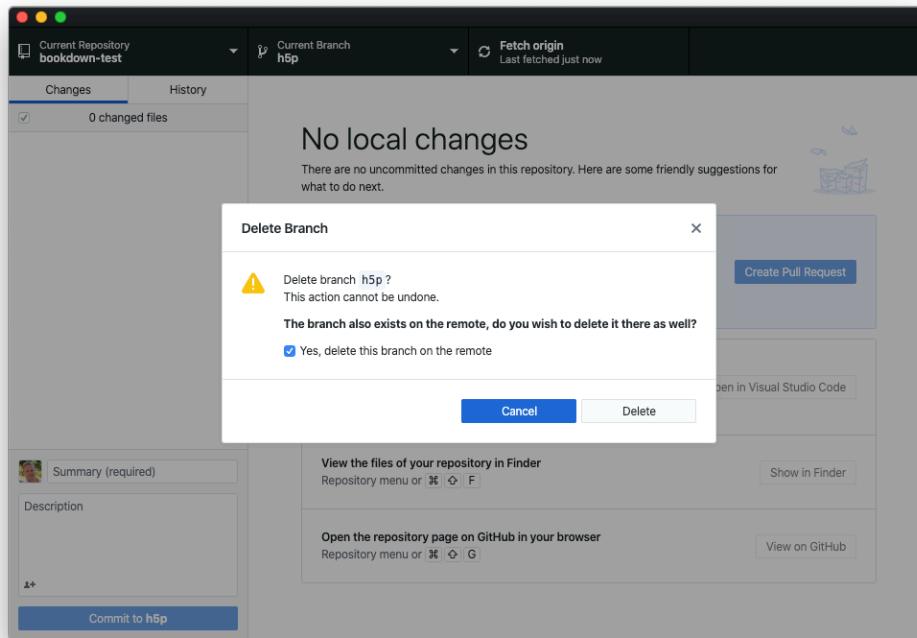


Figure 4.1: La boîte de dialogue d'élimination d'une branche dans GitHub Desktop.

- Changement de gestion de dépôts GitHub suite à l'impossibilité d'utiliser le login et mot de passe depuis septembre 2021. Lancer le shell git et ensuite :

```
git remote -v
# Si on a quelque chose comme: origin https://github.com/org/repo.git c'est pas bon
```

```
# faire:  
git remote set-url origin git@github.com:org/repo.git
```

- Il est utile d'enregistrer le matériel pédagogique tel qu'il se présente chaque année, d'une part à des fins d'archivage, et d'autre part afin que plus tard les étudiants puissent éventuellement retrouver ce matériel tel qu'il était lorsqu'il a suivi le cours. Différentes stratégies sont possibles : utilisation de branches, de releases, ... Cependant pour les bookdowns il n'est possible de présenter le site correspondant que de la dernière version pour un dépôt GitHub donné. Dans ce cas, il est plus intéressant d'archiver le bookdown relatif à une année académique sous forme d'un dépôt séparé, permettant ainsi d'accéder au site web correspondant depuis GitHub Pages. Pour cela, il ne faut pas forkér le dépôt, mais en faire une copie miroir. La procédure pour réaliser cela est expliquée ici. En pratique, nous ajoutons l'année de chaque version au nom du dépôt (par exemple, sdd-umons-2018, sdd-umons-2019, ...). La version correspondant à l'année académique 2018-2019 est sdd-umons-2018 et c'est aussi la version de la SciViews Box associée, la svbox2018. Ensuite, il faut bien entendu éditer le dépôt ainsi copié en miroir pour indiquer dans le préambule que c'est une ancienne version, et aussi remplacer les liens vers des assignations GitHub Classroom obsolètes vers des suggestions de forkér les dépôts templates correspondants afin que les anciens exercices puissent toujours être accédés, mais cette fois-ci, en dehors du contexte de GitHub Classroom.



A partir de l'année académique 2020-2021, le dépôt de travail, ainsi que le lien sur le site web <https://wp.sciviews.org> comportent directement le renseignement de l'année. Ainsi, les modifications nécessaires sont minimales pour archiver le bookdown et la copie en miroir se fait vers le nouveau nom en début d'année.

Chapter 5

Bookdown & learnitdown

Un matériel pédagogique riche et varié est un plus indéniable pour un cours. La même matière présentée de différentes façon stimule les étudiants ayant des sensibilités différentes, et permet une certaine répétition non monotone du message, une des clé d'un apprentissage réussi. Cependant, lorsque nous avons commencé à produire des vidéos, des applications learnr, des dépôts GitHub, des présentations sous forme de diapositives, ... les étudiants étaient un peu perdus dans la quantité de matériel disséminé en différents endroits. **Il est important de centraliser et organiser de manière logique l'ensemble du matériel pédagogique !** Il faut que cette centralisation propose un **file conducteur**, tout en permettant par la suite de **retrouver facilement l'information** souhaitée plus tard. La présentation sous forme de bookdown offre ces atouts :

- En organisant la matière par chapitres, et grâce à la table des matières présente à la gauche, l'étudiant appréhende immédiatement le fil conducteur du cours.
- La table des matières autant que l'outil de recherche dans le bookdown tout entier permet de retrouver plus tard l'information facilement¹.

Le format bookdown est conçu principalement pour écrire des documents de type “text-book” linéaires, même si au fil du temps, il permet d'intégrer des éléments plus interactifs et présentant la matière différemment comme des applications Shiny ou des vidéos. **Les extensions à bookdown proposées par le package {learnitdown}** complète encore les possibilités d'y inclure du matériel interactif et contextuel via H5P, les applications Shiny ou learnr qui peuvent être démarrées *in situ* à la demande, le contenu contextuel en fonction d'une institution ou d'un cours lié à l'étudiant, ou les zones de détail déroulantes. Mais toutes ces additions demandent une adaptation profonde du document bookdown. Le plus facile est donc de démarrer par un fork depuis le dépôt template bookdown-test. La section suivante explique comment créer un nouveau site bookdown agrémenté des ajouts de {learnitdown} via ce fork, et explique ensuite comment l'intégrer dans un site Wordpress. Les fonctionnalités nouvelles offertes par {learnitdown} sont détaillées par après.

¹ Un autre projet nommé **do-it-yourstat** vise à proposer également un tableau de bord complètement personnalisable afin d'organiser son matériel didactique librement et d'y accéder de différentes manières, y compris par cartes mentales, par mots clés, par diagramme décisionnel, ...

5.1 Cration d'un bookdown/learnitdown

Les fonctions supplementaires relatives  H5P et Shiny/learnr proposes par {learnitdown}, ainsi que l'integration  Wordpress impliquent des modifications non triviales dans les paramtres de bookdown. Le plus simple est donc de partir sur la base d'un template fonctionnel. Ce template est fourni  partir du depot GitHub phgrosjean/bookdown-test. La proc re suivante explique comment crer un nouveau bookdown sur base de ce template. Si par contre, l'objectif est de *transformer* un bookdown existant, alors comparez les differents fichiers par rapport  ce meme depot et modifiez-les en consequence.

1. Naviguez vers <https://github.com/phgrosjean/bookdown-test> et cliquez sur le bouton **Use this template**.

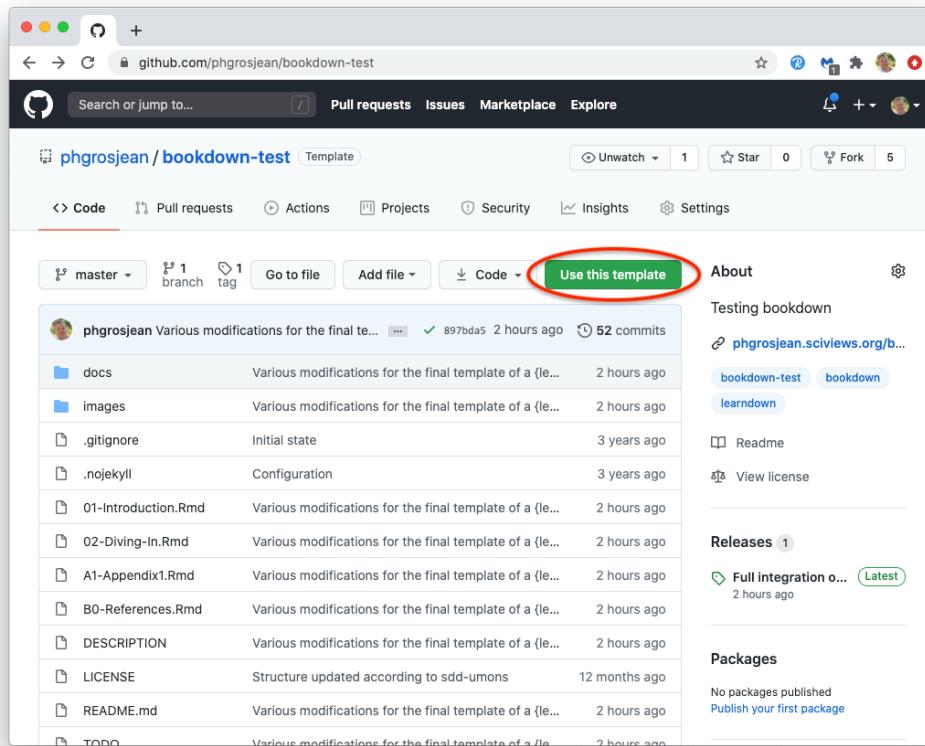


Figure 5.1: Partez du template GitHub `bookdown-test`.

2. Choisissez une organisation pour laquelle vous avez les droits suffisants sur GitHub, et indiquez un nom pour votre nouveau depot. Vous pouvez galement ajouter une description, mettre le depot en priv. Il n'est pas utile d'inclure toutes les branches : seule la branche `master` nous interesse. Ensuite, cliquez sur `Create repository from template`.

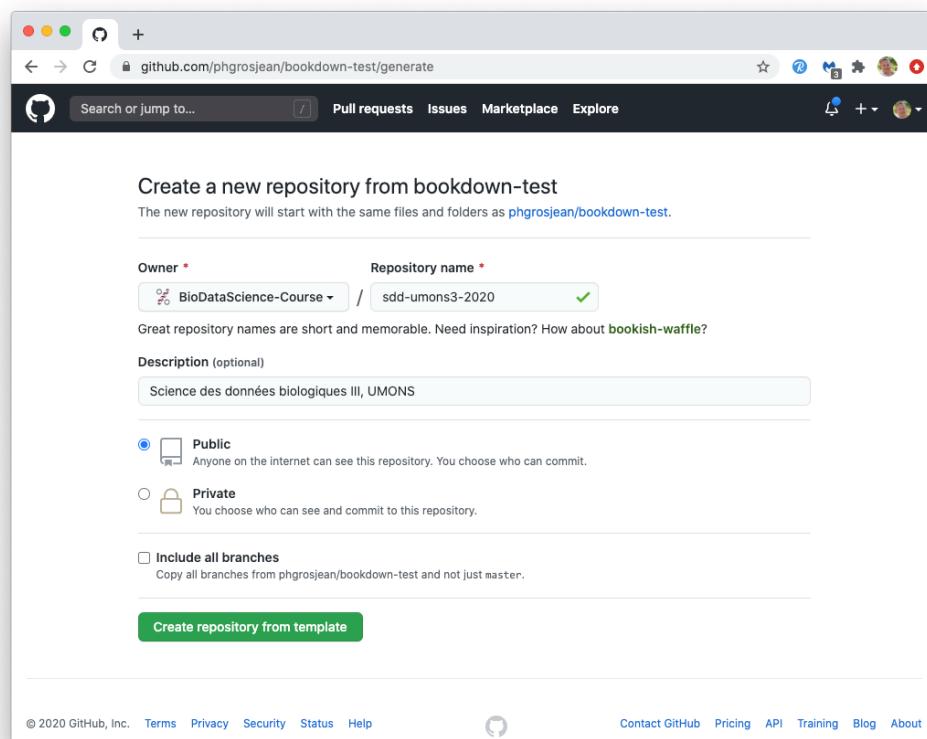


Figure 5.2: Créez un nouveau dépôt.

3. Éditez les options du nouveau dépôt (bouton **Settings** dans la page principale). Le dépôt `bookdown-test` utilisait GitHub-pages pour servir directement le site bookdown à partir du sous-dossier `/docs` mais ici nous serons probablement amenés à inclure les pages du site dans Wordpress. Donc, n'activez donc pas GitHub Pages pour votre nouveau dépôt.
4. Renommez le fichier `bookdown-test.Rproj` en fonction du nom de votre dépôt. Ouvrez-le dans GitHub, puis éditez-le. *Ne changez rien surtout dans le contenu*, mais modifiez son nom au dessus de la boîte d'édition et commitez directement ce changement.
5. Clonez à présent ce nouveau dépôt localement sur votre ordinateur. Vous pouvez le faire à la ligne de commande, via GitHub Desktop, ou directement via RStudio si vous utilisez ce logiciel pour éditer votre bookdown. Dans RStudio, allez dans le menu **File -> New Project...**, choisissez **Version Control** puis **Git** et placez-y l'URL de votre dépôt. Vérifier le dossier dans lequel le dépôt sera placé localement et cliquez sur **Create Project**.

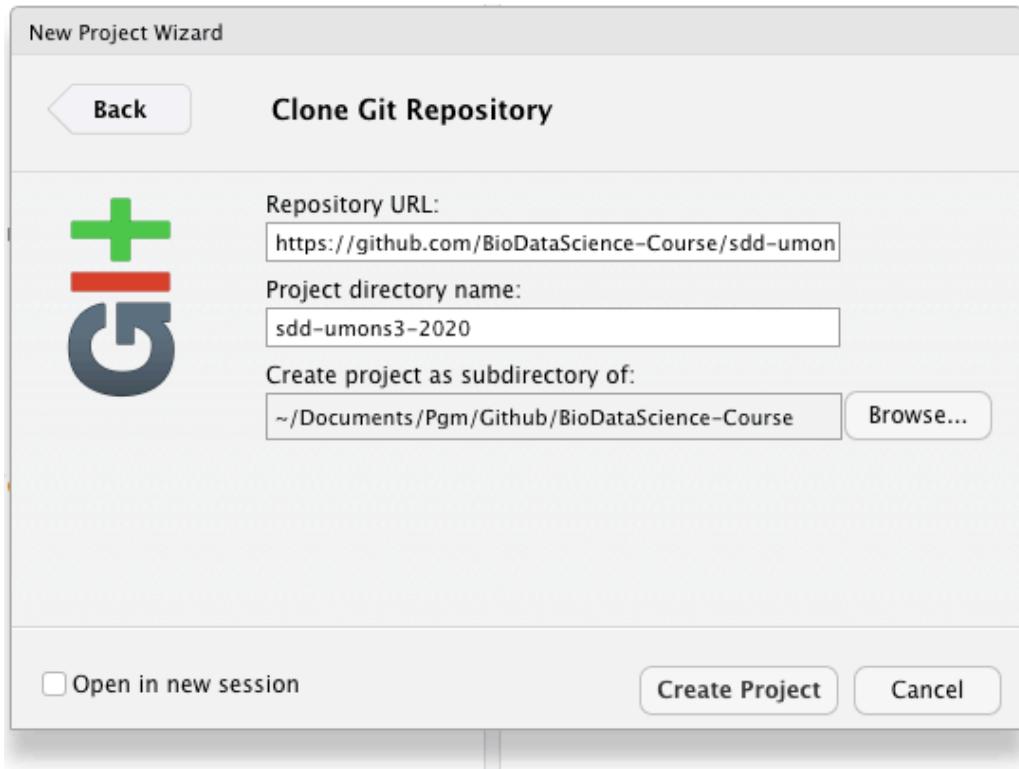


Figure 5.3: Création d'un nouveau projet RStudio à partir d'un dépôt git.

6. Modifiez différents fichiers dans votre nouveau dépôt à partir de RStudio tel qu'expliqué ci-dessous... et éditez votre nouveau bookdown !

5.1.1 README.md

Le premier fichier à modifier est `README.md`. Il doit bien entendu être complètement réécrit en fonction du contexte de votre propre bookdown. Profitez-en pour vérifier la license que vous voulez proposer pour votre oeuvre, indiquez-le dans `README.md` et adaptez éventuellement le fichier `LICENSE` en conséquence.

5.1.2 _bookdown.yml

Le fichier `_bookdown.yml` contient une entrée cruciale pour l'intégration de votre bookdown à Wordpress: `output_dir`. Pour l'instant il est indiqué `output_dir: docs` ce qui signifie que le site bookdown est construit dans le sous-dossier `docs`² Vous allez indiquer un chemin relatif *en dehors du dépôt* (utilisez la notation `..` pour revenir un dossier en arrière). Vous allez pointer vers le sous-dossier `htdocs/nom_bookdown` où `nom_bookdown` est le nom de votre dépôt bookdown dans Wordpress. L'endroit où cela se trouve dépend de la façon dont vous gérez les fichiers de votre site web Wordpress. Dans notre exemple, nous proposons Gandi.net comme hébergeur car il combine nom de domaine, boîtes mails, certificats HTTPS renouvelés automatiquement et site Wordpress complètement préinstallé pour un coût modique. A la section 11.3.0.2 nous avons déjà réalisé un dépôt git associé à notre instance Wordpress. Nous y ajoutons donc le sous-dossier `htdocs`, et allons faire pointer `output_dir` vers le bon emplacement sous forme de chemin relatif. Par exemple dans notre cas, le dépôt bookdown GitHub s'appelle `sdd-umons3-2020` et le dépôt git Gandi.net pour notre Wordpress s'appelle `wp.sciviews.org` et ces deux dossiers sont placés dans le même répertoire. Dans ce cas, nous indiquerons `output_dir: '../wp.sciviews.org/htdocs/sdd-umons3-2020'`.

Profitez-en pour éditer également les autres entrées dans `_bookdown.yml`: `book_filename`, `repo`, `clean` et les labels en fonction de la langue de votre oeuvre.

Une fois que c'est fait et que le fichier `_bookdown.yml` modifié est sauvegardé sur le disque, vous pouvez effacer tout le sous-répertoire `docs` dans le dépôt GitHub de votre bookdown car vous n'en avez plus besoin.

Vérifiez que les fichiers de votre site bookdown vont se placer au bon endroit en le compilant : après avoir installé le package R `{bookdown}`, vous avez un addin `Preview Book` dans le menu déroulant `Addins` de RStudio. Sélectionnez-le et laisser le site se compiler. Ensuite, vérifiez que les fichiers correspondants ont bien été ajoutés dans le sous-dossier `htdocs/nom_bookdown` de votre dossier Wordpress (`wp.sciviews.org` dans notre exemple). Si pas, corrigez le chemin relatif dans `output_dir`.

5.1.3 _output.yml

Ce fichier contient des entrées complémentaires pour paramétriser l'apparence de votre bookdown. Consultez la documentation de bookdown pour vous y aider. Vous voudrez certainement adapter les entrées `config: toc: before:`, `config: toc: after:`, `config: edit:` et `config: view:` en fonction du contexte de votre nouveau dépôt.

²Vous pouvez laisser votre bookdown à cet endroit-là et le servir à l'aide de GitHub Pages si vous voulez, mais alors l'intégration à Wordpress, y compris l'identification des utilisateurs, et l'enregistrement des événements H5P ne seront pas possibles.

5.1.4 index.html

C'est le troisième fichier qui définit le contexte de votre bookdown.

- Adapter `title` et `author`.
- Modifiez `githut-repo` pour pointer vers votre dépôt GitHub.
- Modifiez `l$url` comme celle de votre site Wopress mais avec comme sous dossier `nom_bookdown`. Dans l'exemple, il s'agit de '`https://wp.sciviews.org/sdd-umons3-2020/`'.
- Tant que vous y êtes, adaptez aussi `description`, éditez le contenu de la page et éventuellement aussi, modifiez l'image `front-cover.png` pour refléter le contenu de votre ouvrage.

5.1.5 header.html & header0.html

Le fichier `header.html` est généré automatiquement par `{learnitdown}` et **il ne faut donc pas l'édition**. Si vous voulez rajouter du contenu dans l'entête, incluez-le dans le fichier `header0.html` et ce qui s'y trouve sera placé en bas de `header.html` par `{learnitdown}`. Vous pouvez effacer `header0.html` si vous ne souhaitez rien y rajouter.

5.1.6 footer.html

Placez ici ce que vous désirez en bas de page. Le fichier contient (sous forme commentée), le code nécessaire pour rajouter une zone de discussion Disqus en bas de chaque page. Vous pouvez laisser le fichier en l'état ou l'effacer si vous ne souhaitez rien rajouter en bas de page... ou l'édition en fonction des besoins.

5.1.7 style.css & style0.css

La même logique est d'application pour `style.css` que pour `header.html` (généré automatiquement par `{learnitdown}` et à ne pas édition) et `style0.css` qui contient des styles CSS complémentaires. Le fichier `style0.css` présente un exemple de compléments... à édition en fonction du contexte. Dans cet exemple le style et les couleurs sont légèrement modifiés, et des icônes sont rajoutées pour divers blocs spéciaux. Adaptez ceci à votre propre contexte.

5.1.8 setup.R

Le fichier `setup.R` est utilisé par `{learnitdown}` pour configurer votre bookdown et pour redéfinir les fonctions telles que `launch_shiny()` et `h5p()`. Voyez dans le code et éditez en fonction (des messages en français sont en commentaires).

5.1.9 Contenu

En outre, ce template est prévu pour être utilisé en anglais au départ, mais les entrées qu'il faut édition ont été rajoutées explicitement dans les fichiers de configuration pour vous aider à changer de langue. A noter que les chapitres sont dénommés "modules" et sont supposés correspondre aux différents modules d'apprentissage de votre cours.

Enfin, bien entendu, vous devez modifier le contenu ainsi que les images dans le dossier `images` et ses sous-dossiers. L'image `front-cover.png` est celle affichée en début de document, et les images `block-xxxx.png` sont les icônes affichées pour les différents blocs spéciaux. Les sous-dossiers `intro` et `shinyapps` de `images` contiennent des images relatives au contenu d'exemple dans `01-introduction.Rmd` et peuvent être effacées et/ou remplacées en même temps que ce contenu.



Nous vous conseillons dans un premier temps de *ne pas* modifier la page `01-Introduction.Rmd` ni les sous dossiers `images/intro` et `images/shinyapps` et de terminer la construction et l'intégration de votre site bookdown dans Wordpress. En effet, cette section contient des pages démos qui sont utiles pour tester si les différentes fonctionnalités sont opérationnelles dans votre propre configuration. Ce n'est qu'après avoir effectué ces tests en inspectant le contenu de ce chapitre d'introduction à partir de l'URL définitive que vous pouvez alors remplacer cette page par votre propre contenu.

5.2 Intégration dans Wordpress

Le format gitbook de `{bookdown}` qui est le format le plus utilisé pour les tutoriels, les cours et les manuels (type “textbook”) en ligne n'est pas prévu pour s'intégrer à Wordpress. Cependant, grâce à `{learnitdown}`, il est possible d'obtenir un résultat indiscernable visuellement d'un réelle intégration via l'inclusion du bookdown dans un iframe à l'intérieur d'une page Wordpress. De plus, `{learnitdown}` transfère des informations relatives à l'utilisateur Wordpress enregistré, et aux fonctions côté serveur pour l'enregistrement des données H5P. Ainsi, l'intégration est complète puisque les applications H5P incluses dans le bookdown bénéficient des mêmes fonctionnalités que si elles étaient directement affichées à l'intérieur d'une page Wordpress. Les mêmes fonctions de détection de l'utilisateur et d'enregistrement des évènements générés par les applications Shiny et learnr sont également implémentées dans `{learnitdown}`. La partie à modifier dans le bookdown pour réaliser cette intégration ayant été abordée ci-dessus, voyons maintenant ce que nous devons faire du côté du serveur Wordpress pour la finaliser.

1. Inclusion des pages du bookdown dans le serveur Wordpress

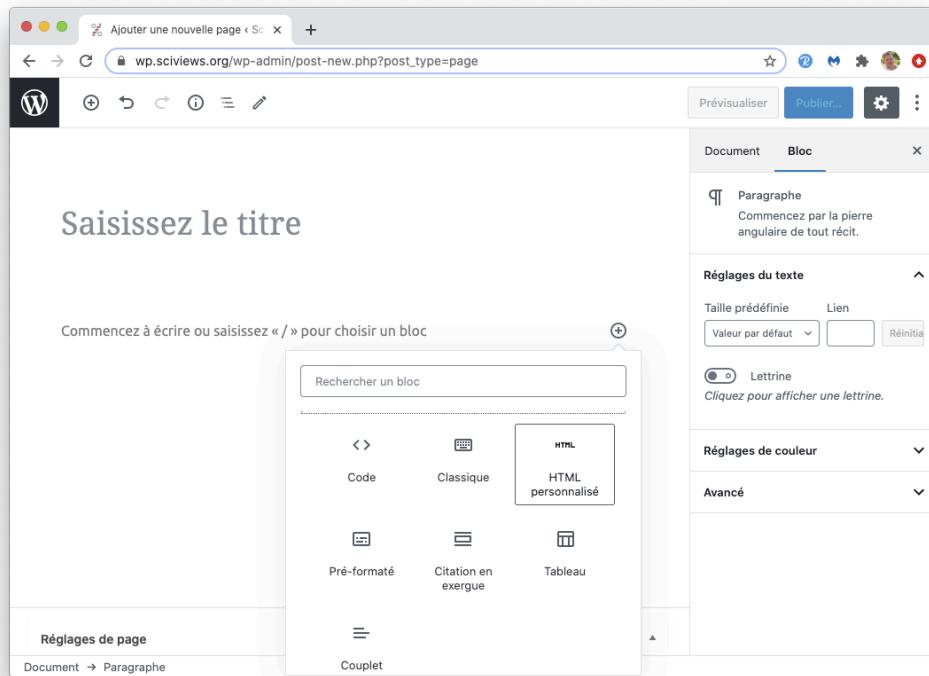
Les pages du bookdown ont été générées dans un sous-dossier dans le dépôt git `wp.sciviews.org` dans notre exemple. Nous devons maintenant effectuer un commit, un push sur Gandi.net, et enfin un deploy. Ceci peut se faire dans un premier temps à la ligne de commande pour test :

```
cd <path_to_wp.sciviews.org>
git add .
git commit -m "New version of sdd-umons3-2020"
git push origin master
ssh 3693753@git.sd3.gpaas.net 'deploy wp.sciviews.org.git'
```

TODO: Automatiser ceci, plus utiliser correctement l'instruction `clean` de Gandi.net.

2. Créer une page contenant uniquement un iframe dans Wordpress

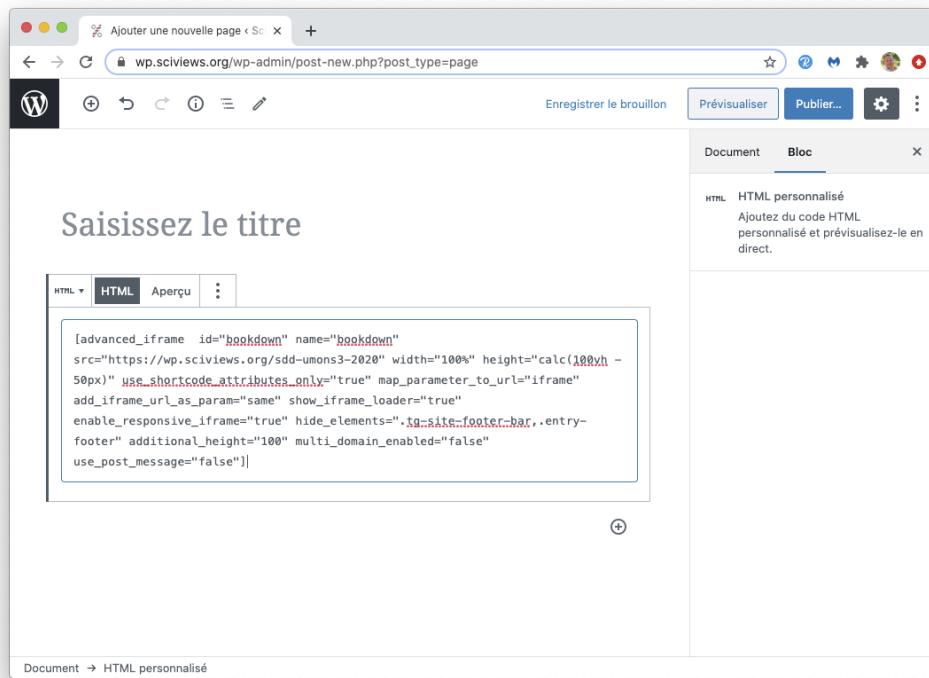
Avec un compte ayant des droits en édition, aller dans le tableau de bord de Wordpress, ensuite dans **Pages-> Ajouter**. Dans l'éditeur, n'indiquez aucun titre, mais ajouter un bloc en dessous et sélectionner le type **HTML personnalisé**.



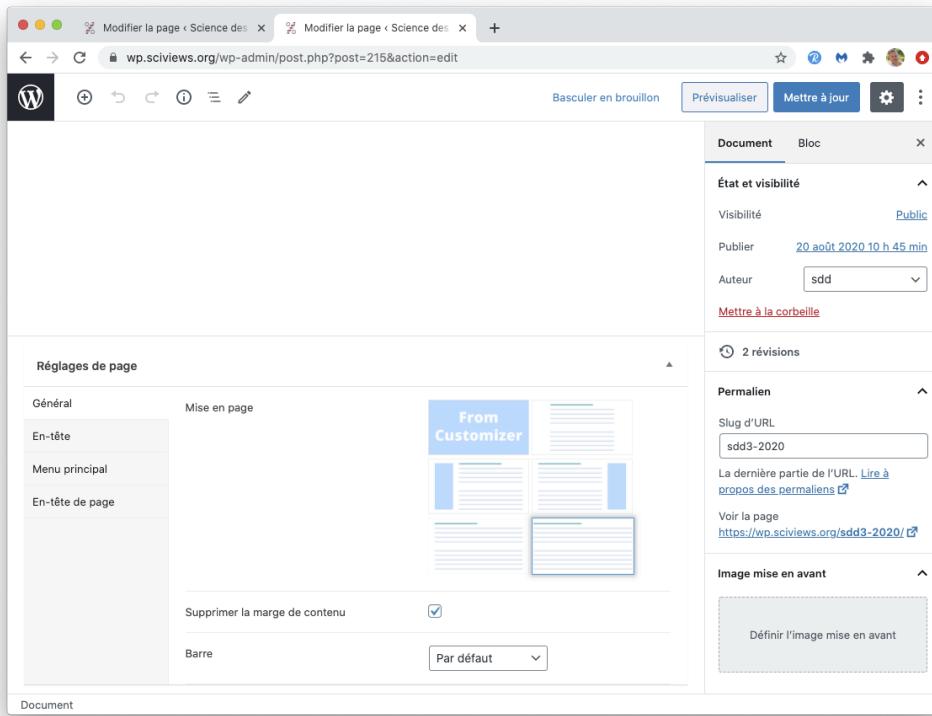
Collez-y le code ci-dessous et adaptez le contenu de `src` en fonction de votre propre page, ainsi que la hauteur de la barre supérieure en pixels dans `height="calc(100vh - 50px)"` (ici, 50 pixels).

```
[advanced_iframe id="bookdown" name="bookdown" src="https://wp.sciviews.org/sdd-umons3-2020"
width="100%" height="calc(100vh - 50px)" use_shortcode_attributes_only="true"
map_parameter_to_url="iframe" add_iframe_url_as_param="same" show_iframe_loader="true"
enable_responsive_iframe="true" hide_elements=".tg-site-footer-bar,.entry-footer"
additional_height="100" multi_domain_enabled="false" use_post_message="false"]
```

Appuyez sur le bouton **Publier**. Vous devez obtenir quelque chose qui se rapproche de ceci :



3. Allez dans l'onglet Document, puis Permalien-> Slug d'URL et choisissez un nom de la page *qui diffère de celui du dépôt GitHub* (donc dans notre cas, pas sdd-umons3-2020). Par exemple, sdd3-2020.



4. Enfin, allez tout en bas de l'éditeur dans **Réglages de la page -> Général** et sélectionnez la mise en page la plus simple en bas à droite et cochez également **Supprimer la marge du contenu**.
5. Cliquez sur le bouton **Mettre à jour**. Visualisez votre page et naviguez dans la section **Introduction** du bookdown pour tester les différentes fonctionnalités.
6. A partir d'un compte administrateur Wordpress, cliquez sur **Personnaliser** dans la barre noire supérieure. Ensuite, éditez le menu qui apparaît à gauche pour y inclure un lien vers la page que nous venons de créer.

Au final, le bookdown se présente sous deux formes :

- Les pages affichées telles quelles via une URL reprenant le nom du dépôt GitHub (dans notre cas, il s'agit de <https://wp.sciviews.org/sdd-umons3-2020/>). Nous avons une présentation bookdown classique. Les informations de l'identité de l'étudiant sont transmises via `localStorage`, mais le bandeau supérieur du site et la barre noire supérieure de gestion Wordpress ne sont pas affichés. Le bookdown n'est pas vraiment intégré dans le site et les événements xAPI de H5P ne sont pas renvoyés au serveur Wordpress. Cette représentation est utile pour l'utilisation du bookdown dans un contexte hors cours UMONS.
- Les mêmes pages du bookdown, **mais intégrées sous forme d'iframe** dans une page de Wordpress dédiée à partir du lien de la page Wordpress (dans notre exemple, <https://wp.sciviews.org/sdd3-2020>). La page conserve des éléments issus de Wordpress comme la barre d'outils supérieure (accès à Moodle, Discord, etc.) ainsi

que la barre noire supérieure de contrôle Wordpress qui permet aussi de vérifier d'un coup d'œil que l'on est connecté, et sous quelle identité nous le sommes. S'il veut lire le contenu du bookdown sans voir la bannière haute, le lecteur peut faire coulisser la page jusque tout en bas et cette bannière viendra se cacher derrière la barre noire supérieure Wordpress (qui, elle, reste toujours visible tant que l'utilisateur est connecté au site).

5.3 Contenu H5P dans bookdown

Nous avons déjà vu que le contenu H5P s'affiche dans un iframe dans un document principal. Il lui transmet les événements xAPI, et c'est donc ce document principal qui doit les traiter. Nous avons ici une impossibilité à le faire si nous sommes dans un cas où H5P est servi par un serveur différent que notre bookdown. En effet, les accès en “Cross-Origin Resource Sharing” ou CORS sont restreintes pour éviter des attaques de site évidentes. Ainsi un iframe issu d'un serveur différent de la page qui l'héberge ne peut communiquer librement avec la page qui le contient si elle provient d'un site différent. Or, nous avons besoin de cette communication pour transmettre les événements xAPI...

La solution que nous avons trouvée est d'ajouter le dossier qui contient nos pages statiques du bookdown à Wordpress directement. Ainsi les deux sont maintenant servis à partir de la même adresse <https://wp.sciviews.org>. Une fois que c'est fait, l'inclusion d'applications H5P dans le bookdown est facile. Il suffit d'utiliser la fonction `h5p()` de `{learnitdown}` en indiquant juste le numéro ID de l'application H5P tel qu'affiché dans le tableau de bord de Wordpress avec un code du type `learnitdown::h5p(6, baseurl = "https://wp.sciviews.org")` dans du code inline dans notre bookdown (ici, l'ID serait ‘6’, ce qui correspond à l'application H5P nommée “drag-words” dans la copie d'écran ci-dessous). On peut éventuellement repréciser `width` ou `height` = pour changer la taille de l'iframe hébergeant l'application H5P.

Pour comprendre ce que la fonction `learnitdown::h5p()` fait, nous pouvons nous concentrer sur la version simplifiée suivante :

```
h5p <- function(id, width = 780, height = 500) {
  glue::glue("\n<iframe src=\"https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id={id}\">
```

La fonction construit un fragment HTML qui ajoute un iframe et un script pour gérer l'inclusion du contenu H5P dans notre bookdown. A partir d'ici, ajouter un contenu H5P est aussi simple que d'ajouter `h5p(id)` dans du code R en ligne où “id” est l'ID que nous avons noté. On peut aussi jouer sur les arguments `width=` et `height=` pour ajuster l'emplacement et la taille du contenu H5P dans la page.

Ainsi programmé notre contenu H5P se comporte comme suit :

- pour un utilisateur enregistré dans Wordpress, les événements xAPI sont enregistrés à son nom. Nous pouvons donc suivre la progression de nos étudiants via les contenus H5P parsemés dans le bookdown également, en plus des applis Shiny et des learnrs. De plus, l'état d'achèvement de l'exercice est aussi enregistré pour un bon nombre de contenus H5P et rétabli lorsque la page est rechargée par le même utilisateur.

The screenshot shows the WordPress admin dashboard for the 'Science des données biologiques' site. The left sidebar includes links for Articles, Médias, Pages, Commentaires, Appearance, Extensions, Utilisateurs, Outils, Réglages, miniOrange OAuth, and WP-Optimize. The main content area is titled 'All H5P Content' and displays a table of five entries:

Title	Content type	Author	Tags	Last modified	ID	Results	Edit
drag-words	Drag Text	sdd	test	2020/05/19	6	Results	Edit
multiple-choice	Multiple Choice	sdd	test	2020/05/19	7	Results	Edit
R markdown	Interactive Video	sdd	sdd1, sdd1-02	2020/05/19	8	Results	Edit
Description de la SciViews Box	Drag Text	sdd	sdd1, sdd1-02	2020/05/19	9	Results	Edit
Le nuage de points : Visualiser des données numériques avec R & SciViews::R	Interactive Video	sdd	sdd1, sdd1-02	2020/05/19	10	Results	Edit

At the bottom of the page, there is a footer note: 'Merci de faire de WordPress votre outil de création.' and 'Version 5.4.1'.

Figure 5.4: H5P Content dans Wordpress. Les widgets sont identifiés par leur **ID** (ici, 6, 7, 8, 9 & 10).

- pour un utilisateur non enregistré, les contenus H5P fonctionnent normalement, mais aucun événement xAPI n'est enregistré, et l'état de complétude de l'exercice n'est pas conservé non plus. Si la page est rechargée, les contenus H5P retournent à leur état initial à chaque fois.

Ceci est un inconfort mineur pour un utilisateur anonyme et permet l'utilisation des contenus H5P de manière optimale pour nos étudiants dûment enregistrés, tout en permettant une libre utilisation par d'autres visiteurs sans polluer notre base de données.

Nous avons également tenté de passer des arguments supplémentaires via l'URL qui lance la page bookdown, et qui les transfère vers l'iframe H5P, mais sans succès car apparemment Wopress réécrit l'URL et laisse tomber ces arguments supplémentaires. Mais de toutes façons, ce n'est pas important pour l'instant car nous pouvons tracer nos utilisateurs enregistrés dans Wordpress telle que c'est actuellement implémenté.

Chapter 6

Tutoriels learnrs

TODO: place learnr stuff here...

Chapter 7

Applications interactive Shiny

...

Chapter 8

Exercices H5P

H5P permet de “créer, partager et réutiliser du contenu interactif HTML5 dans un browser Web”, d’après la page d’introduction du site Web. Il s’agit d’un ensemble de contenus HTML5 (html + css + javascript) permettant d’afficher et d’interagir avec des exercices et des présentations sous plusieurs formes. Plus de 40 contenus différents sont disponibles. H5P est distribué sous forme de logiciel libre (gratuit et Open Source) qu’il est possible d’installer sur son propre serveur. De plus, h5p.org propose l’hébergement de contenus gratuits **à des fins de test uniquement**, et offre en complément une formule payante sur h5p.com, mais le tarif n’est pas renseigné.

Dans cette section, nous reprendrons les choix techniques, l’installation, la configuration et le déploiement de H5P dans le cadre de nos cours de science des données à l’UMONS, ainsi que les principaux contenus H5P qui sont disponibles.

8.1 Installation de H5P

H5P ne peut fonctionner seul. Il doit être inclus dans un serveur plus large, soit un LMS (Learning Management System ou plateforme d’apprentissage en ligne comme), soit un CMS (Content Management System, système de gestion de contenu).

L’un des LMS les plus populaires est Moodle qui a été choisi à l’UMONS comme outil institutionnel d’apprentissage en ligne. H5P est compatible avec Moodle (il peut être installé comme plugin). Ce serait donc parfaitement logique de l’utiliser de cette manière. Un test avait été fait en 2019. Si cela fonctionne, nous avons pu constater que le contenu H5P n’est **pas** utilisable ailleurs que dans Moodle lui-même, en tous cas de la façon dont le serveur test était configuré, avec les sécurités nécessaires pour protéger le site. Or, notre souhait est d’intégrer le contenu H5P dans un bookdown qui ne peut être hébergé sur Moodle¹. L’essai qui a été réalisé à l’UMONS en 2019 n’a pas été concrétisé dans le serveur Moodle en production. Or, il nous fallait intégrer du contenu H5P dans nos bookdown, et ce, dès 2020-2021. Donc, nous avons recherché une autre option.

Cette autre option consiste à utiliser soit Wordpress, soit Drupal pour lesquels H5P est également compatible. Après comparaison des possibilités des deux, nous avons opté pour

¹D’un point de vue technique, le contenu H5P s’affiche dans un iframe. Or les possibilités des iframes inter-sites sont limitées par le protocole HTML5 sous-jacent lui-même pour éviter les abus. Il nous faut donc le bookdown (pages statiques) et le contenu H5P (devant être lié à un serveur LMS ou CMS) servis par le *même* site.

Wordpress, plus largement utilisé.

H5P permet de récupérer des informations concernant l'activité des étudiants dans les exercices sous forme de requêtes xAPI. xAPI est un standard développé pour permettre l'interaction entre un LMS comme Moodle qui produit des évènements xAPI, et un LRS (Learning Record Store) une base de données chargée d'enregistrer les traces de l'apprentissage au format standardisé xAPI. En complément de H5P lui-même et de Wordpress comme serveur principal, il nous faut donc un LRS et un plugin qui fasse le pont entre H5P et ce LRS. Enfin, l'information n'est utile que si elle est identifiable comme provenant d'un étudiant en particulier. Le serveur doit donc gérer l'authentification des utilisateurs et transmettre l'identité de l'étudiant au travers de la requête xAPI. Si tout cela paraît très compliqué, voici un petit schéma qui résume l'ensemble (à noter que nous avons choisi l'authentification via GitHub parce que nos étudiants ont également d'autres activités sur GitHub/GitHub Classroom, et il est dès lors plus facile de gérer une identité commune à toutes ces activités, même si l'identité UMONS est, hors contexte, la plus logique ici) :

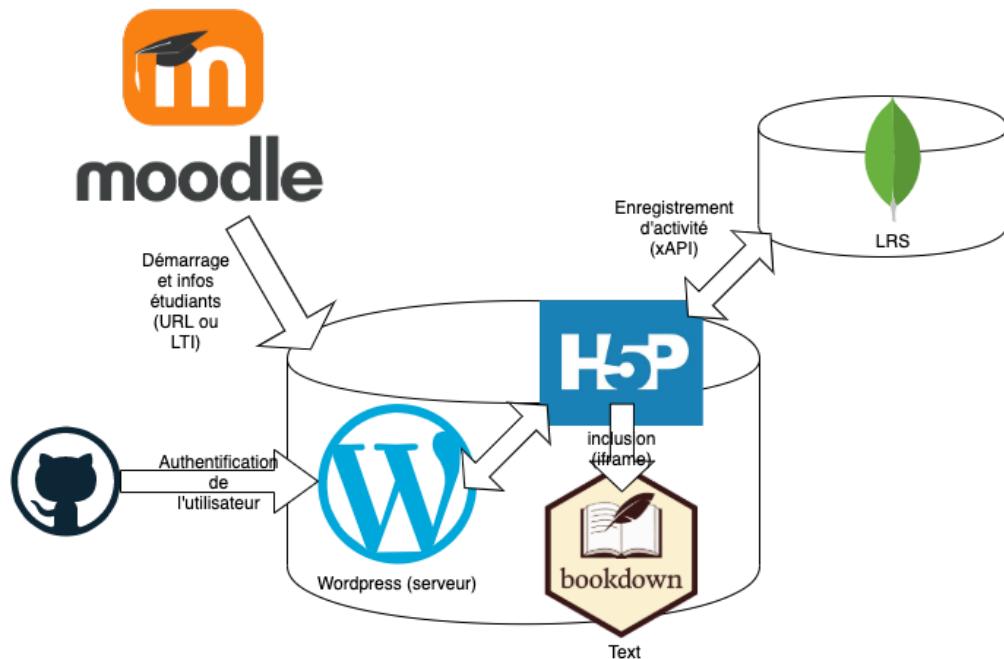


Figure 8.1: H5P et bookdown sont intégrés dans le même serveur Wordpress, ce qui permet l'affichage en iframe sans contraintes. Les informations de l'utilisateur proviennent d'une authentification GitHub, plus des données issues de Moodle via URL ou LTI. L'activité des étudiants H5P peut ainsi être enregistrée dans le LRS de notre choix via le protocole standard xAPI avec une identification univoque.

8.1.1 Installation du plugin H5P

Une fois le serveur Wordpress opérationnel, l'installation de H5P lui-même est vraiment très simple (voir aussi la page d'installation Wordpress de H5P).

- Connection au serveur Wordpress en temps qu'administrateur,

- Aller dans le tableau de bord -> Extensions -> Ajouter.
- Dans le champ de recherche, introduire “H5P”,
- Sélectionner “Interactive content - H5P”. Cliquer “Installer”, puis “Activer”.

C'est fait ! H5P est installé, mais aucune forme de contenu n'est encore disponible. H5P conseille d'installer des **librairies** de contenu. A partir du tableau de bord, cela se fait via “H5P Content -> Libraries”. Cependant, cela n'apparaît pas nécessaire en pratique car lorsqu'on importe du contenu H5P dans le site, les librairies nécessaires sont automatiquement importées également... plus d'explications ci-dessous.

8.1.2 Configuration de H5P

Le panneau de configuration est accessible pour l'administrateur Wordpress depuis le tableau de bord -> Réglages -> H5P. Voici une copie d'écran de la configuration choisie.

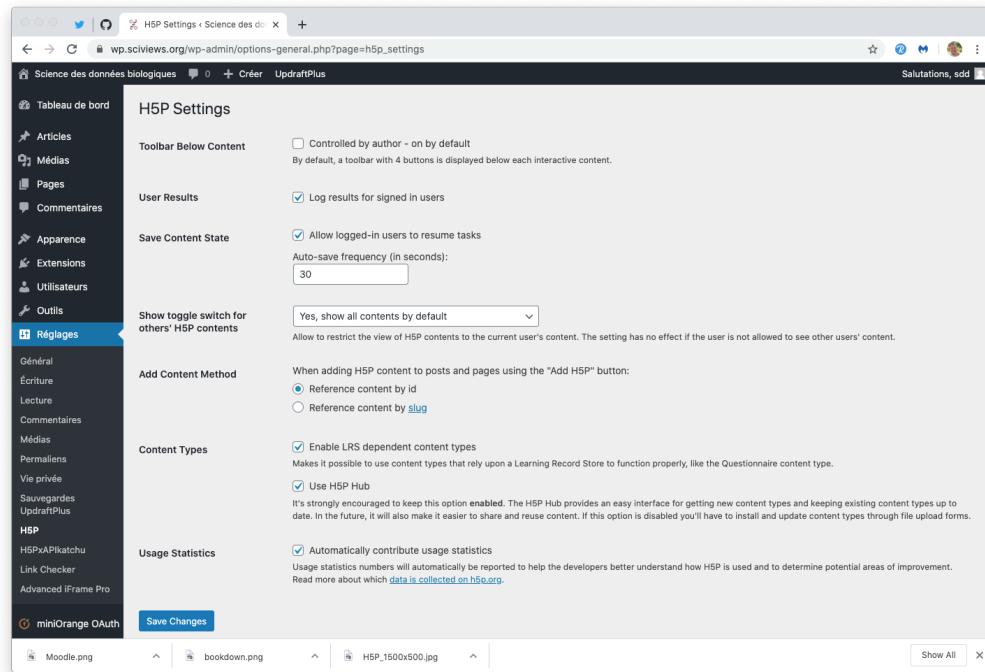


Figure 8.2: Tableau de configuration du plugin H5P dans Wordpress.

- La barre d'outils H5P en dessous du contenu est désactivée par défaut (“Controlled by author” non coché). Cette barre d'outils permet le partage et la réutilisation de contenu. Nous la garderions bien activées, mais le bouton de réemploi doit apparemment être activé pour que l'inclusion dans un bookdown soit possible. Or, si ce bouton est activé, tout le monde peut inclure le contenu dans d'autres sites... et impacter notre quotas d'utilisation de requêtes. Nous ne le souhaitons bien évidemment pas. *Pour récupérer un contenu sous forme de fichier .h5p, il faut réactiver cette option temporairement dans le tableau de bord, et seul l'administrateur du site peut le faire (voir Gestion des contenus H5P plus loin).*

- Les résultats utilisateurs (“Log results...” est coché) permet d’avoir des statistiques de base utilisateur par utilisateur sur les H5P. C’est utile, mais il faudra voir à terme l’impact sur l’espace disque utilisé pour cette fonction.
- La sauvegarde du contenu (“Allow logged-in users to resume tasks”) est activée et utile puisqu’elle permet la mémorisation d’une session à l’autre des sélections dans les contenus H5P pour les utilisateurs enregistrés.
- “Show toggle switch for others’ H5P contents” n’est pas très clair pour moi. Je l’ai laissé pour l’instant sur “Yes, show all contents by default”. Est-ce qu’il s’agit de voir le contenu *créé* par d’autres utilisateurs ? Si, oui, c’est ce qu’on veut (les étudiants utilisent le contenu créé par les profs).
- “Add content method” est réglé sur “id”.
- L’option suivante, “Enable LRS dependent content types” est importante et doit être cochée pour pouvoir enregistrer l’activité via xAPI (voir plus loin).
- Les deux dernières options sont en relation avec la gestion et le développement H5P. Nous les laissons cochées (“Use H5P Hub” et “Automatically contribute usage statistics”).

8.2 Enregistrement H5P xAPI

Une fois le plugin H5P installé, il nous faut encore permettre l’enregistrement des activités des étudiants. Nous l’avons déjà vu, cette possibilité est prévue dans H5P et le protocole standardisé xAPI est utilisé, ce qui le rend compatible avec les LRS du marché. Dans Wordpress, nous avons deux options :

1. Un enregistrement *local* dans la base de données MySQL de Wordpress directement sur le serveur avec le plugin Wordpress **H5PxAPIkatchu**, voire une modification de ce plugin pour l’enregistrement dans une base de données généraliste, comme MongoDB,
2. Un enregistrement dans un LRS spécialisé xAPI avec le plugin Wordpress **wp-h5p-xapi**.

Nous allons aborder et discuter ces deux options tour-à-tour.

8.2.1 H5PxAPIkatchu

Ce plugin est facile à installer puisqu’il est disponible depuis Wordpress plugins. La procédure est la même que pour l’installation de H5P lui-même (l’administrateur va dans la partie installation d’extension du tableau de bord, rentre “H5PxAPI” dans la zone de recherche, sélectionne l’item et clique “Installer” puis “Activer”). Ce plugin est gratuit et il semble bien maintenu².

La configuration est également extrêmement simple.

Comme nous souhaitons récolter un maximum d’information sur l’activité des étudiants, et voulons récupérer le contenu xAPI intact, nous cochons “Store complete statements” et “Capture everything”. L’option “Debug” n’est pas nécessaire et peut rester décochée.

²J’ai fait une proposition de pull request à l’auteur qui a répondu immédiatement.

The screenshot shows the 'H5PxAPIkatchu' configuration page under the 'Réglages' section. It includes sections for 'General' settings and 'H5P content types'. The 'General' section has three checkboxes: 'Store complete statements' (checked), 'Debug' (unchecked), and 'Capture everything' (checked). The 'H5P content types' section displays a table with five rows of data:

	Title	Type	Id
1	Définition de Github	Multiple Choice	11
2	Le nuage de points : Visualiser des données numériques avec R & SciViews::R	Interactive Video	10
3	Description de la SciViews Box	Drag Text	9
4	R markdown	Interactive Video	8
5	multiple-choice	Multiple Choice	7

Figure 8.3: Configuration de H5PxAPIkatchu.

Une fois cette configuration effectuée, l’activité H5P est enregistrée directement dans la base de données MySQL de Wordpress. Cette base de données n’est pas disponible de l’extérieur dans notre configuration (le fournisseur verrouille les bases de données pour un accès local uniquement par sécurité). Par contre, H5PxAPIkatchu permet de visualiser un tableau avec les données, de le filtrer, et de récupérer le résultat sous forme de fichier CSV avec le bouton “Download” (le lien “Delete” permet d’effacer le contenu et n’est disponible que pour l’administrateur du site).

Et c’est tout ! Pas besoin de LRS, pas besoin de configuration à rallonge... par contre, cela ne convient que pour du *dépannage*. En effet, l’accès aux données n’est pas possible pour, par exemple, générer les rapports de progression des étudiants. L’extraction sous forme CSV doit se faire *manuellement* !

Après l’enthousiasme suite à un premier essai d’enregistrement de l’activité réussie lorsque le contenu H5P est utilisé directement dans Wordpress (à partir d’une page Wordpress ou du tableau de bord), nous faisons face à une déception quand nous constatons que cela ne **fonctionne pas dans le bookdown** ! Pourtant, nous avons bien installé notre bookdown dans le *même* serveur que Wordpress/H5P pour permettre une communication fluide... alors, pourquoi ça ne fonctionne toujours pas ? En fait, nous devons nous plonger dans les détails de l’implémentation pour comprendre que ce n’est **pas** H5P lui-même qui effectue l’enregistrement, mais la page dans laquelle le contenu est inclus. Autrement dit, notre brave bookdown en pages statiques va se voir déléguer le travail. Le schéma suivant explique le détail de l’implémentation.

Moralité, pour que notre bookdown enregistre les événements H5P, il faut lui adjoindre le code en Javascript nécessaire pour faire le transfert. Ce code est repris et adapté de H5P

Actor Id	Actor Name	Actor Group Members	Verb Id	Verb Display	Object Id	Object Def. Name	URL
Général	Philippe views.org Grosjean	Lecture	http://adonet.gov/expapi/verbs/attempted	attempted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=8?subContentId=3f31e62e-9c9c-4571-9a7d-ad08060f3588	Untitled Single Choice Set	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=8?subContentId=3f31e62e-9c9c-4571-9a7d-ad08060f3588
Écriture	Philippe views.org Grosjean		http://adonet.gov/expapi/verbs/answered	answered	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=6	drag-words Box	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=6
Lecture	Philippe views.org Grosjean		http://adonet.gov/expapi/verbs/interacted	interacted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=6	drag-words Box	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=6

Figure 8.4: Contenu xAPI enregistré par H5PxAPIkatchu. L’utilisateur est correctement identifié dans le tableau enregistré.

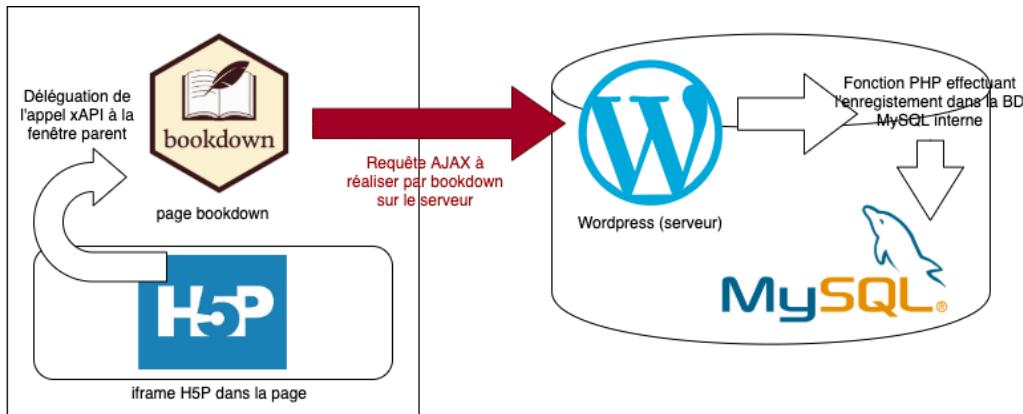


Figure 8.5: Pour des raisons de sécurité, l’enregistrement de l’événement xAPI se fait “server-side” par du code PHP dans H5PxAPIkatchu. Par contre, H5P ne communique jamais lui-même avec le LRS, mais délègue cette tâche à la page parent qui l’héberge (ce qui ne fonctionne que si les deux sont issus du même domaine). Donc, notre bookdown va devoir effectuer le transfert de l’événement xAPI via une requête AJAX sur le serveur (partie en rouge) qui manque actuellement.

lui-même, ainsi que de H5PxAPIkatchu. Ce code est inclus dans le fichier `header.html` qui est inclus dans le bookdown grâce à ceci dans `index.Rmd` :

```
output:
  bookdown::gitbook:
    includes:
      in_header: header.html
```

Tout d'abord, la partie H5P de configuration. La partie **H5PIntegration** est normalement générée à partir de Wordpress, mais comme nous utilisons des pages statiques dans bookdown, nous le copions tel quel ici. Comme elle contient des informations contextuelles liées au site et à l'installation de H5P, nous devrons récupérer cela autrement afin d'avoir une version à jour, mais pour l'instant, ça fonctionne comme cela. Voici ce que cela donne :

```
<!-- Required for H5P xAPI feedback -->

<script>H5PIntegration = {"baseUrl": "https://wp.sciviews.org", "url": "\/wp-content\/uploads\/h5p-integration.js", "css": ["h5p-core-styles-h5p-confirmation-dialog-css", "h5p-core-styles-h5p-core-button-css"], "js": ["wp-embed.min.js"]}</script>
```

Et c'est pas tout ! Il nous faut encore le code Javascript H5P et H5PxAPIkatchu :

```
<script>
var wpAJAXurl = "https://wp.sciviews.org/wp-admin/admin-ajax.php";
var debugEnabled = "0";
var captureAllH5pContentTypes = "1";
var h5pContentTypes = [""];
window.H5PxAPIkatchu = {
  captureAllH5pContentTypes: captureAllH5pContentTypes,
  debugEnabled: debugEnabled,
  h5pContentTypes: h5pContentTypes,
  jQuery: jQuery,
  wpAJAXurl: wpAJAXurl
};
</script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/jquery.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-event-dispatcher.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-x-api-event.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-x-api.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-content-type.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-confirmation-dialog.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/h5p-action-bar.js'></script>
<script src='https://wp.sciviews.org/wp-content/plugins/h5p/h5p-php-library/js/request-queue.js'></script>
```

Voilà, avec tout cela, nous recompilons le bookdown, plaçons tout le contenu de \docs sur le serveur... et (après bien des essais et erreurs tout de même)... ça marche !



En dépit de sa simplicité, H5PxAPIkatchu ne convient *pas* tel quel pour l'objectif que nous poursuivons. Nous avons imaginé étendre le plugin pour enregistrer les données dans notre base de données MongoDB Atlas comme première solution. Cela a nécessité du développement PHP et l'utilisation d'une librairie PHP mongodb qu'il faut installer sur le serveur. Une autre solution consiste à utiliser un autre plugin et à passer par un véritable LRS. Ceci est envisagé dans le paragraphe suivant.

Il s'avère que, une fois l'incorporation du driver mongodb dans PHP effectuée (c'est en fait la partie la plus difficile), le code en PHP qu'il faut ajouter à H5PxAPIkatchu pour effectuer l'enregistrement dans notre base MongoDB est assez succinct. Une demande à l'auteur de H5PxAPIkatchu pour permettre l'enregistrement des données ailleurs via une fonction PHP a été rapidement honorée. Ainsi, nous avons pu facilement écrire le code complémentaire pour enregistrer l'information dans notre base de données MongoDB ATLAS (ce code est placé dans la zone dédiée du plugin **My custom functions**) :

```
// Custom h5pxapikatchu data insertion
function h5pxapi_actor( $actor ) {
    global $doc;
    $doc['actor'] = $actor;
    return $actor;
}
add_filter( 'h5pxapikatchu_insert_data_actor', 'h5pxapi_actor' );

function h5pxapi_verb( $verb ) {
    global $doc;
    $doc['verb'] = $verb;
    return $verb;
}
add_filter( 'h5pxapikatchu_insert_data_verb', 'h5pxapi_verb' );

function h5pxapi_object( $object ) {
    global $doc;
    $doc['object'] = $object;
    return $object;
}
add_filter( 'h5pxapikatchu_insert_data_object', 'h5pxapi_object' );

function h5pxapi_result( $result ) {
    global $doc;
    $doc['result'] = $result;
    // Added by PhG
    // Note: 'time' is like recorded in MySQL, but truncated down to second
    // $doc['time'] = current_time( 'mysql' );
    $doc['date'] = current_time( 'Y-m-d H:i:s.u', $gmt = 1 );
    return $result;
}
```

```

add_filter( 'h5pxapikatchu_insert_data_result', 'h5pxapi_result' );

function h5pxapi_xapi( $xapi ) {
    global $doc;
    $doc['xapi'] = $xapi;
    return $xapi;
}
add_filter( 'h5pxapikatchu_insert_data_xapi', 'h5pxapi_xapi' );

//function h5pxapikatchu_custom_insert_data( $actor, $verb, $object, $result, $xapi ) {
function h5pxapi_custom_insert_data() {
    global $doc;
    $response = false;

    try {
        $manager = new MongoDB\Driver\Manager('mongodb://sdd-wordpress:Pseud0chr0mis-pacagnellae@');
        $bulk = new MongoDB\Driver\BulkWrite;
        // $doc = [
        //   'actor' => $actor,
        //   'verb' => $verb,
        //   'object' => $object,
        //   'result' => $result,
        //   // Added by PhG: 'time' => $time,
        //   'xapi' => $xapi,
        // ];
        $bulk->insert( $doc );
        $result = $manager->executeBulkWrite( 'sdd.h5p', $bulk );
        $response = true;
    } catch (Throwable $e) {
        //echo "Catured Throwable: " . $e->getMessage() . PHP_EOL;
        // Manage error here...
    }
    if ($response) {
        wp_die();
    }
    return $response;
}
add_action( 'h5pxapikatchu_insert_data_pre_database', 'h5pxapi_custom_insert_data' );

```

Cela fonctionne, mais pour une utilisation plus robuste, il nous faudra passer les paramètres nécessaires à la requête AJAX sur Wordpress de manière dynamique à nos pages bookdown. C'est ce que nous avons implémenté dans le package R `{learnitdown}`, voir chapitre 5.

8.2.2 wp-h5p-xapi et LRS

L'autre plugin Wordpress qui permet d'enregistrer les événements xAPI est `wp-h5p-xapi`. Dans cette configuration, il nous faut un LRS. La plupart des solutions proposées sont commerciales, payantes et ... chères, sauf une : Watershed. Son utilisation comme LRS est

gratuite et sans limitations ni du nombre d'utilisateurs, ni du nombre d'événements générés par mois. Ça semble presque incroyable, sauf si on comprend que c'est pour eux un produit d'appel vers leurs outils analytics qui effectuent des rapports sur le LRS, et qui sont eux, bien chers (minimum 1600\$/mois) ! Mais nous, du moment qu'on puisse enregistrer nos événements xAPI et les récupérer dans R, on a tout ce qu'il nous faut. Donc, la solution complète envisagée passerait par wp-h5p-xapi et Watershed.

En allant sur leur site, on se rend compte qu'il n'est pas possible de créer un compte directement. Il faut envoyer une demande via un formulaire... et la réponse se fait attendre... indéfiniment ! De plus, autre déconvenue : le plugin Wordpress wp-h5p-xapi n'a pas reçu de commit depuis plus de 2 ans (7 avril 2018) et semble ne plus être conforme aux spécifications de Wordpress (voir ici) de sorte qu'il n'est plus distribué par le site Wordpress plugins depuis le 31 octobre 2019 en raison de violation des lignes directrices de Wordpress... Ce n'est pas très encourageant.

8.2.3 Autres plugins H5P

- Le même auteur de H5PxAPIkatchu propose aussi H5P Resize Pulse qui permet de régler un problème d'affichage apparemment de H5P dans certains containers. Dans les bookdowns, nous ne rencontrons pas de problèmes pour l'instant, mais c'est bon à savoir... Il propose aussi H5P sharing pour obtenir aisément le code `<iframe>...</iframe>` d'un contenu H5P. Nous n'en avons pas besoin car la fonction `learnitdown::h5p()` remplit la même fonction dans R.
- H5P CSS Editor adapte le look de l'éditeur H5P à celui du site Wordpress, ... mais ce plugin n'a pas été mis à jour depuis plus de 3 ans. Nous ne l'avons pas testé.
- Gamipress est une solution Open Source et entièrement gratuite de gamification de sites Wordpress (l'utilisateur est encouragé à utiliser le site en gagnant des points, des badges, un ranking des utilisateurs peut être affiché, etc.). Ils proposent aussi un plugin qui permet d'obtenir des points avec les activités H5P. Les badges et la gamification de l'apprentissage est une piste très intéressante que nous devrions développer pour stimuler nos étudiants. Donc, nous explorerons cette piste plus tard (voir aussi la section sur les badges dans le chapitre 12).

8.2.4 Solution finale choisie

Nous avons vu l'extrême complexité à faire fonctionner H5P/Wordpress avec bookdown. Cependant, nos développements ont abouti à une solution fonctionnelle via le plugin H5PxAPIkatchu paché, puis complété par son auteur. Nous avons donc décidé de finaliser cette implémentation de la façon suivante :

- Suite à un pull request, l'auteur de H5PxAPIkatchu a modifié son plugin pour permettre de faire appel à une fonction personnalisée pour enregistrer les événements xAPI de H5P où bon nous semble. Nous l'utilisons via une fonction écrite en PHP dans Wordpress pour enregistrer ces informations dans MongoDB ATLAS, comme expliqué plus haut.
- Toute la machinerie Javascript nécessaire dans la page qui inclus un composant H5P dans un iframe est fournie au bookdown via l'extension dans le package R `{learnitdown}`. Ceci est discuté dans la section correspondante (voir chapitre 5).

8.3 Gestion des contenus H5P

Chaque contenu H5P est indépendant du reste. Attention que les contenus complexes sont volumineux. Par exemple, les vidéos interactives occupent chacune 2,8Mo sur le serveur lorsque les vidéos sont sur youtube (donc, occupation de place minimale puisque la vidéo elle-même n'est **pas** sur le site Wordpress).

Au départ, aucun contenu n'est réalisable car les **librairies** nécessaires ne sont pas encore installée. Mais il n'est pas nécessaire de les installer car il suffit de téléverser du contenu pour que leur installation se fasse automatiquement (mais si de nouvelles librairies sont nécessaires, il faut alors le faire avec le compte admin de Wordpress). Donc, la procédure pragmatique suivante est retenue pour enrichir le site des librairies dont nous avons besoin :

- Pour un type de contenu qui n'est pas encore supporté, aller sur h5p.org. Se connecter (créer un compte gratuit, si nécessaire).
- Aller dans "Examples and Downloads" et choisir un exemple de type de contenu souhaité. Cliquer sur "Reuse". Ou alors aller dans "Contents" -> "Create New Content" et créer un contenu du type souhaité. Puis cliquer sur "Reuse" lorsqu'on est satisfait du contenu.

Voici un exemple avec le contenu de type "True/False".

Ensuite, dans le tableau de bord de Wordpress (en ayant pris soin de s'être enregistré comme administrateur), aller dans H5P Content -> Add New. Sélectionner "Téléverser" et indiquer le fichier .h5p que l'on vient de sauver, puis "Utiliser".

Une fois que c'est fait, non seulement le contenu est téléchargé, mais les librairies requises sont également ajoutées. De cette façon, un autre contenu de ce type peut maintenant être créé *directement* depuis le tableau de bord de Wordpress, H5P content -> Add New par n'importe quel utilisateur qui a les droits d'édition sur le site (donc, les enseignants).

8.3.1 Backup H5P

Le backup du site entier Wordpress est effectué. Par contre, il est utile d'avoir une copie de nos différents contenus H5P au format de fichier .h5p. Pour cela, un administrateur de site doit au préalable temporairement cocher l'option "Toolbar Below Content" dans les paramètres du plugin H5P de Wordpress (tableau de bord), de sorte que le bouton "Reuse" apparaisse en dessus d'un contenu H5P. Pour chaque contenu que l'on souhaite sauvegarder, l'afficher, et cliquer sur son bouton "Reuse" pour récupérer le fichier .h5p. Ensuite, placer les fichiers .h5p sur p-Cloud dans le répertoire `sdd_h5p_backup`.

TODO: une solution de récupération automatisée des fichiers .h5p d'un site devrait être développée à la place, la procédure décrite ci-dessus ne devrait être utilisée que ponctuellement pour récupérer un ou un petit nombre de contenus H5P.

8.4 Quelques contenus H5P utiles

H5P propose plus d'une quarantaine de types de contenus différents. Tous ne sont pas utiles dans le contexte des cours de science des données, mais certains le sont particulièrement. Voici une petite sélection.

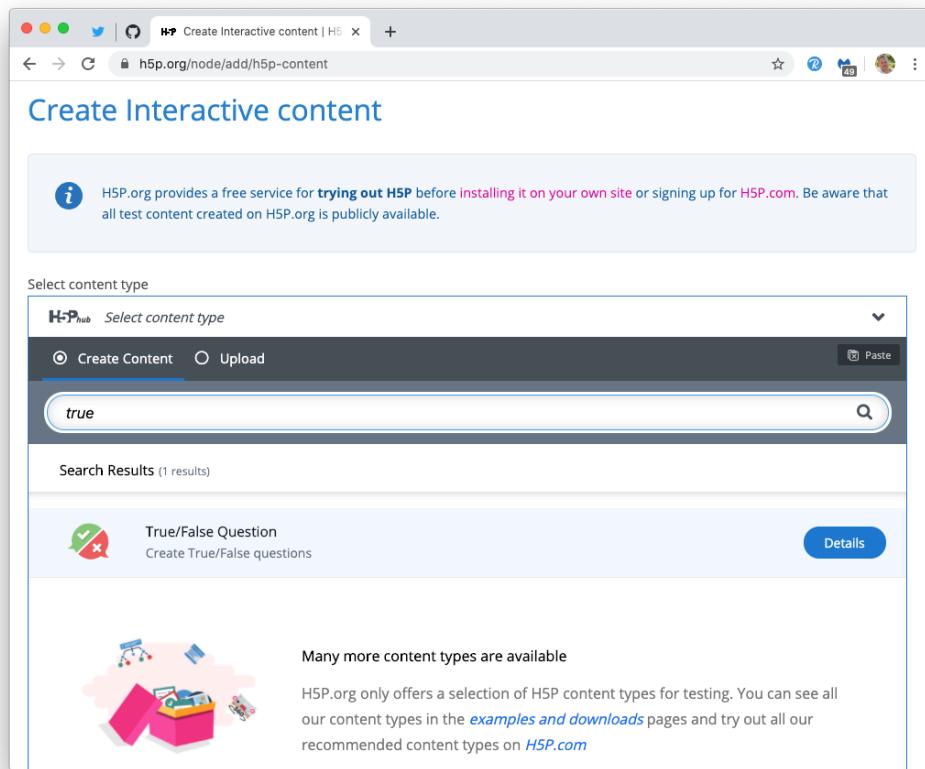


Figure 8.6: Dans son compte sur h5p.org aller dans “Contents” -> “Create New Content”, et choisir le type de contenu que l’on veut créer (ici True/False question).

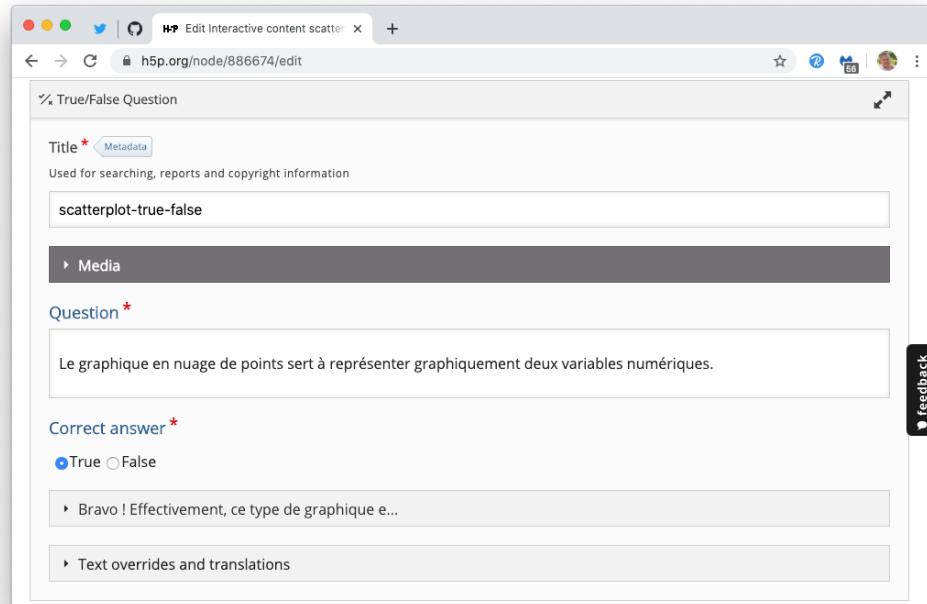


Figure 8.7: Créer le contenu et l'enregister.



Figure 8.8: Une fois le contenu terminé, il suffit de cliquer sur “Reuse” (flèche rouge) et de cliquer sur “Download as an .h5p file” pour récupérer le fichier .h5p.

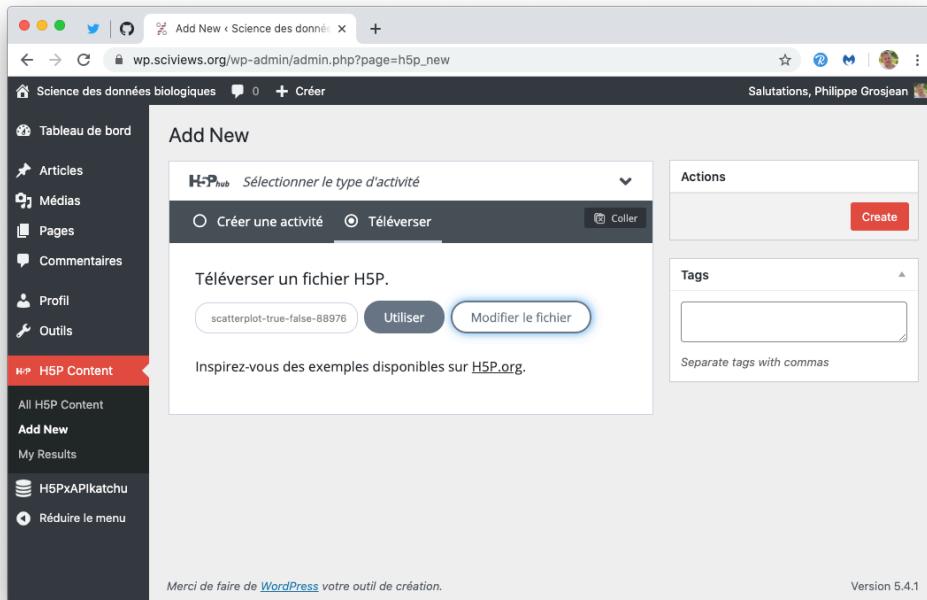


Figure 8.9: Ajout du contenu H5P sur le site Wopress par téléversement du fichier .h5p.

8.4.1 Apprentissage du code R

Apprendre à coder en R, comme tout apprentissage de langage, nécessite du temps et de la persévérance. L'approche pas à pas, ludique et diversifiée est, en la matière, importante. H5P nous aide notamment avec les contenus suivants :

- Advanced fill the blanks. Il est plus facile au début de se focaliser sur des parties du code lorsqu'un exemple est décortiqué. Cependant, l'approche purement passive (l'étudiant suit une explication sur le code) n'est pas très fructueuse. L'étudiant doit être **actif** et forcé à **réfléchir**, tout en l'amenant en douceur à écrire *lui-même* des (parties) de code. “Advanced fill the blanks” est parfait pour cela car il permet un feedback efficace pour aider l'étudiant dans son apprentissage. Ce genre de question pourrait également être envisagée dans une zone de code R d'un learnr, mais c'est plus lourd et à réserver à une phase ultérieure de l'apprentissage.
- S'il faut encore plus faciliter l'exercice, nous pourrions aussi suggérer des fragments de code à replacer au bon endroit, c'est ce que fait Drag the words. C'est en quelque sorte une version ultra-simplifiée de l'exercice de type “remplir les vides” puisque les solutions sont aussi suggérées sous forme de fragments de code à replacer au bon endroit.
- Pour varier, nous pourrions aussi proposer des bouts de code et poser ensuite une question simple sur ce code, à répondre par oui ou non, voir True/False question.
- Pour l'apprentissage de la syntaxe R (où sont les variables, les fonctions, les arguments, les chaînes de caractères, etc. dans du code), nous pouvons utiliser Mark the words.

8.4.2 Apprentissage des graphiques

La visualisation des données impose de développer deux compétences en parallèle : (1) la faculté de déterminer quelle est la meilleure représentation graphique pour quelles données, et (2) la capacité à maîtriser techniquement l'outil pour effectivement réaliser le graphique voulu. En plus de cela, il faut aussi que le graphique soit correct (libellé ou unité des axes présents, respect des conventions, pas de surcharges inutiles, etc.)

- Pour montrer les différentes parties importantes d'un graphique ou, par exemple, expliquer les différents éléments d'une boîte de dispersion, l'Image hotspots est idéal. C'est infiniment plus engageant qu'un graphique et une légende statiques.
- Le complément logique pour faire des exercices d'auto-évaluation sur le même thème, c'est le Find multiple hotspots.
- Le Drag and drop propose une alternative intéressante pour varier le type d'exercices consistant à identifier les parties importantes d'un graphique.
- Image juxtaposition permet de comparer deux solutions graphiques différentes d'une manière originale. Il est aussi possible de visualiser une variation sur un thème lié à un même graphique, par exemple, deux palettes de couleurs différentes, ou l'utilisation de la couleur *versus* des symboles. Nous pouvons aussi imaginer avec cet outil comparer un modèle linéaire et un modèle polynomial, par exemple.

8.4.3 Explications générales

- Ici, les Vidéos interactives sont un must. Les étudiants étant immergés dans l'audiovisuel constamment, l'image animée est un canal de communication familier et habituel. Nous devons l'exploiter. Par contre, il est trop facile de regarder passivement une vidéo... et d'avoir l'*impression* d'avoir compris sans pourtant n'avoir rien assimilé. Parsemer des questions sur ce qui vient d'être visualisé dans la vidéo est un moyen extrêmement puissant pour faire prendre conscience que la lecture est passive. L'étudiant peut alors revenir en arrière et visionner plus attentivement dans le but de pouvoir répondre à la question. Ce faisant, il apprend, comprend et assimile bien mieux le contenu de la vidéo !
- Il ne faut pas négliger non plus le Course presentation. Si les explications deviennent longues et les pages de texte interminables, ce n'est pas très engageant pour l'étudiant. La matière sous forme d'une présentation où des questions d'auto-évaluation sont directement incluses est un moyen très efficace de diversifier, casser la monotonie d'une longue explication textuelle, et plus engageant si la présentation est richement illustrée. Pour un exemple concret, si nous expliquons comment réaliser un test t de Student et en même temps que nous donnons les détails de calcul, cela fait trop. De plus, de part notre expérience, l'étudiant a ensuite du mal à faire la part des choses entre l'application du test dans R et les détails plus techniques du "comment ça marche en interne". Le texte devient aussi très (trop) long. Une solution est alors de se focaliser sur la résolution du test dans R uniquement sous forme de développement textuel, et réserver une présentation "comment le t test est calculé" pour les détails, et peut-être une autre présentation intitulée "les différents t tests" pour expliquer les différentes variantes.

8.4.4 Autoévaluation

Les learnrs sont des outils de choix que nous appliquons avec succès pour l'autoévaluation des étudiants. Cependant, une autoévaluation plus rapide et plus proche du matériel pédagogique lui-même est encore plus efficace en première ligne (suivi en seconde ligne par des learnrs avec un niveau de difficulté supérieur, en troisième ligne des exercices guidés sur données réelles dans un projet RStudio *via* GitHub Classroom et enfin en quatrième ligne, des analyses à faire *via* GitHub Classroom mais avec un guidage minimal). Nous venons de voir que H5P permet cette autoévaluation directement à l'intérieur des vidéos et des présentations. Il nous manque l'équivalent pour les développements textuels. Les contenus suivants fournissent ce dont nous avons besoin à ce niveau.

- Quizzes permet de rassembler plusieurs questions successives en un contenu unique avec une note finale pour l'ensemble du genre 4/6 questions répondues correctement.
- Pour chaque question du quiz, voire des questions isolées parsemées dans le texte directement, nous avons différents contenus possibles : vrai/faux, choix unique ou Multiple choice.
- Pour une variation plus ludique des questions d'auto-évaluation, nous pouvons envisager utiliser les Flashcards, surtout pour les graphiques.

8.4.5 “Take home message”

Dans une capsule d'apprentissage certains éléments sont cruciaux. Bien définir les objectifs et les prérequis au début de la capsule sont vitaux. En fin de capsule, un résumé des points importants à retenir, le fameux “take home message” est ultra-important également... encore faut-il que l'étudiant assimile et mémorise ce/ces message(s). H5P propose du contenu qui permet d'appréhender ce résumé de manière plus active et ludique.

- Summary se *doit* pratiquement d'être utilisé systématiquement en fin de chaque capsule d'apprentissage pour résumer les points importants. Avec ce contenu H5P, c'est l'étudiant qui établi sa liste de messages importants de manière **active**.

Chapter 9

Projets GitHub Classroom

TODO: décrire ici la gestion des projets GitHub et {learnitgrid}.

Part II

LearnIt::R LRS

Chapter 10

Learning Record Store & MongoDB

Un “Learning Record Store” (LRS en abbrégé) est une base de données un peu particulière configurée pour stocker toutes les données relatives au système d’apprentissage dans un format spécifique nommé xAPI. Ce dernier format permet d’interconnecter différents systèmes qui produisent ou qui consomment des données entre eux. Il existe de nombreux outils payants Learning Locker, Watershed LRS, ou gratuits tel SCORM Cloud LRS ou TRAX LRS pour gérer ce type de base de données. Ces outils assurent la compatibilité nécessaire.

Note : il existe aussi un autre (ancien) standard en plus de xAPI nommé SCORM. Une initiative cmi5 vise à fusionner les deux et à fournir une interface standardisée pour interopérer des outils différents au sein d’un “Learning Management System” (LMS) hétérogène. En gros, cmi5 = xAPI + LMS, voir P.27 ici.

10.1 Notre “LRS”

Les évènements que nous enregistrons ne sont pas tous formatés en xAPI pour le moment. Ainsi, les données provenant de H5P le sont, mais pas celles provenant de Learnr, Shiny, ou Wooclap ne le sont pas. Nous avons décidé d’un format commun de table qui regroupe toutes ces données dans une base de données MongoDB. Plus tard, ce format pourra éventuellement évoluer vers un autre qui soit conforme à xAPI et qui puisse être utilisé dans un véritable LRS.

Le choix de MongoDB a été motivé par plusieurs facteurs :

- La possibilité d’utiliser gratuitement MongoDB ATLAS, un serveur sécurisé et fiable contre les pannes sur le cloud pour y stocker (temporairement) jusqu’à 512Mo de données.
- La facilité d’installer un serveur MongoDB local uniquement (machine personnelle et RStudio Connect).
- La simplicité de backup/restauration de données pour un même serveur ou entre serveurs avec `mongodump/mongorestore`.

- La facilité d'exportation des données avec `mongoexport`, ainsi que les drivers pour PHP (Wordpress/H5PxAPIKatchu) et pour R avec le package {mongolite}.
- La rapidité de requêtes relativement complexes, et la possibilité de mettre au point de telles requêtes avec MongoDB Compass.

Par contre, au niveau des points négatifs, nous avons :

- Utilisation d'un port particulier 27017 qui doit être ouvert. À l'UMONS ce port est fermé pour certains réseaux (Wifi étudiant) car il est aussi contenu dans une plage utilisée par des jeux. Or, il est nécessaire de limiter l'accès aux jeux sur le réseau du campus car sinon, cela saturerait la bande passante.
- Pas de possibilité d'utiliser un outil comme {dbplyr} pour définir les requêtes. Langage de requêtage spécifique.
- Plus complexe à déployer et utiliser en local que SQLite.
- Des solutions plus performantes et mieux intégrées à R et Python existent comme Spark au RAPIDS/BlazingSQL.
- Il faut utiliser un format de données différent pour l'échange et la publication des données (Open Data).

Nous ne sommes donc pas définitivement fixés sur cette solution, mais elle a le mérite de fonctionner correctement jusqu'ici.

10.1.1 Vue générale

- Création d'un compte MongoDB ATLAS, une organisation UMONS et un serveur M0 SSD-COURS. Deux bases de données sont créées : "test" et "sdd". La première ne contient rien d'important et sert... à des tests comme son nom l'indique. La seconde contient nos données avec les collections "h5p", "learnr" et "shiny".
- Création d'un serveur local uniquement sur sdd.umons.ac.be qui contient la base de données "sdd", et des tables "users", "h5p", "learnr", "shiny", ... Également, mise en place de scripts cron (voir plus loin) pour récupérer les données venant de MongoDB ATLAS toutes les 5min et pour faire des backups journaliers de la base de données.
- Utilisation d'un serveur MongoDB local sur une machine sous Windows ou MacOS avec un snapshot des données obtenu via `mongodump/mongorestore`. Ce snapshot, ainsi que les programmes nécessaires pour l'utiliser sous MacOS et Windows est partagé via p-Cloud.

Ces trois systèmes sont détaillés ci-dessous.

10.1.2 MongoDB ATLAS

TODO: expliquer l'organisation de MongoDB ATLAS...

10.1.3 Serveur MongoDB local associé à RSConnect

TODO: installation et configuration dans sdd.umons.ac.be

10.1.4 Serveur local MacOS ou Windows

Le principe de cet item est de regrouper dans un dossier tous les programmes et toutes les données relatives à un snapshot donné de notre LRS, afin de pouvoir manipuler tout cela de manière indépendante et sécurisée (accès local uniquement). Le dossier de base contient deux sous-dossiers : **bin** et **db**.

Le dossier **bin** contient tous les exécutables MacOS et Windows nécessaires, soit **mongod** et **mongo** de la distribution 4.4.11 de MongoDB, et **bsondump**, **mongodump**, **mongoexport**, **mongofiles**, **mongoimport**, **mongorestore**, **mongostat** et **mongotop** des MongoDB Data Tools version 100.5.1. Ces versions sont choisies identiques à celles déployées sur sdd.umons.ac.be et la version majeure de MongoDB est alignée sur celle déployées sur MongoDB ATLAS¹. Ces différents exécutables sont récupérés depuis les installateurs (format .tgz ou .zip) MacOS et Windows depuis MongoDB on-premises et MongoDB tools. L'ensemble de ces exécutables MacOS et Windows occupent un peu plus de 1Go dans le dossier **bin**.

Le dossier **db** contient le snapshot de la base de données. Il est créé en réalisant un **mongorestore** à partir des données obtenues par **mongodump** depuis sdd.umons.ac.be comme ceci (exemple sous MacOS) :

```
# On sdd.umons.ac.be, go to a emty folder, say '/data1/dump/sdd', then:
mongodump # Possibly restrict with --collection=<col_name> and/or --query='{"x": {"$gt":1}}'
# Compress this folder
cd ..
tar -czvf sdd.tar.gz sdd
rm -rf sdd
# Go back to your local machine, then:
cd <empty_temporary_dir>
scp econum@sdd.umons.ac.be:/data1/dump/sdd.tar.gz sdd.tar.gz # Enter password...
tar -xzvf sdd.tar.gz
rm sdd.tar.gz
cd ..
<path_to>/mongorestore # Of course, the local MongoDB server must be running here
```

En plus de ces deux sous-dossier, il y a également deux scripts qui permettent de lancer le serveur MongoDB sous MacOS ou sous Windows. Ces scripts sont :

- **sdd_server_macos.sh** :

```
#!/bin/bash
cd "$(dirname "$0")"
port=27017
./bin/mongod --port $port --dbpath db
```

¹ Aligner les versions majeurs de MongoDB avec MongoDB ATLAS nécessite de jongler un peu avec les mises à jour car nous n'avons aucun contrôle des versions sur ATLAS. Par exemple, une migration de 4.4 vers 5.0 a été planifiée début février 2022 et nous devons adapter tout le reste au même moment !

```

• sdd_server_win.bat

@echo off
set port=27017
start cmd /c bin\mongod --port %port% --dbpath db

```

Ces deux scripts sont à exécuter dans un terminal MacOS ou une fenêtre Powershell Windows après avoir changé le répertoire par défaut vers le dossier racine :

```

cd <dossier_racine>
./sdd_server_macos.sh ou ./sdd_server_win.bat

```

TODO: Un script qui remplit le sous-dossier `bin` en fonction du système, et qui traite aussi le cas Linux Ubuntu.

Et ainsi un démon MongoDB serveur local est lancé sur le port 127.0.0.1:27017 et sert les données présentes dans le sous-dossier `db`. Les messages du serveur sont directement affichés dans la fenêtre terminal. À partir de ce moment, on peut travailler tranquillement sur les données, les analyser et/ou prototyper des outils à déployer ensuite sur `sdd.umons.ac.be` sans craintes de casser la base de données principale. Ce serveur ne permet qu'un accès local. Donc, il est parfaitement sécurisé.

Note : Les versions du serveur et des outils doivent toujours être les mêmes que celles sur MongoDB ATLAS et `sdd.umons.ac.be`. En cas de changement de version, il faut donc bien penser à remplacer les exécutables du dossier `bin` par les nouvelles versions avant de continuer à travailler !

Pour échanger ce snapshot et ces outils entre ordinateurs, le dossier racine et son contenu peuvent être compressés (zippé) et partagés via le dossier **Crypto Folder** de p-Cloud. **Ne pas exécuter le serveur MongoDB directement depuis le dossier p-Cloud pour des raisons évidentes de performances**, mais copier le fichier et le dézipper sur un disque dur rapide qui contient suffisamment de place (5Go au moins).

10.1.5 Mise-à-jour de MongoDB

La version de MongoDB serveur est dictée par celle installé dans MongoDB ATLAS. Par exemple, en février 2022, la version change de 4.4 à 5.0 et il faut adapter tous les outils en fonctions. Voici les étapes à réaliser :

- Vérification de la compatibilité du driver PHP dans Wordpress et upgrade éventuel (voir la section consacrée à Wordpress). Ici, il a fallu passer de la version 1.6 à la version 1.9.
- Vérification de la compatibilité de {mongolite}. La version installée avec la SciViews Box 2021 est la 2.3.1, et il faut au moins la 2.4.1 donc dans R :

```

packageVersion("mongolite")
# If needed:
#remotes::install_github("jeroen/mongolite@v2.4.1")

```

- Vérification de l'option “featureCompatibilityVersion” qui doit être mise à 4.4 (nous voulons pour l'instant rester compatible avec l'ancien sans utiliser les fonctions nouvelles de MongoDB 5.0), donc dans R :

```
library(mongolite)
# Start the local server, then... check featureCompatibilityVersion (4.4)
admin <- mongo(url = "mongodb://127.0.0.1:27017/admin")
admin$run('{ "getParameter": 1, "featureCompatibilityVersion": 1 }')
# If it is lower than 4.4, check first upgrade recommandations, then
#admin$run('{ "setFeatureCompatibilityVersion": "4.4" }')
```

- Mettre à jour les versions de MongoDB Tools. Il nous faut la version 100.5.1 qui est déjà celle installée dans sdd.umons.ac.be. Faire cet upgrade sur les versions locales aussi.
- Tester la compatibilité MongoDB 5.0 sur une version locale de la base de données et y exécuter les différents outils.
- Le jour où MongoDB ATLAS passe sur MongoDB 5.0, installer les exécutables MongoDB 5.0 dans sdd.umons.ac.be et les versions locales de MongoDB partout.

10.2 Learning Analytics Platform

Le stockage des données est une première étape. Il faut ensuite avoir des outils qui digèrent, analysent et présentent ces données sous des formes utiles pour l'enseignant. Ce sont les “Learning Analytics Platforms” (LAP). Nous développons de tels outils pour notre plateforme en utilisant R et {Shiny}. TODO: suite...

10.3 Mise à jour de MongoDB

Notre système utilise deux serveurs MongoDB différents (et même trois, si on considère utiliser une version locale également dans la phase d'analyse des données). Un de ces serveurs est MongoDB ATLAS sous forme gratuite (et limitée à une taille de 512Mo). Ce serveur sert à collecter temporairement les données provenant des machines virtuelles des étudiants (learnrs et applications Shiny exécutées en local), ainsi que les données provenant des exercices H5P.

L'accès à ces bases de données se fait via deux drivers différents : mongodb-php et {mongolite} pour R. De plus, le backup et le transfert des données d'une base à l'autre se fait via mongodump et mongorestore.

Les versions des différents serveurs aussi bien que les drivers et outils doivent tous être compatibles entre eux, naturellement. Or, dans sa version gratuite, MongoDB ATLAS ne laisse pas le choix de la version, ni du moment où elle est mise à jour. Toutes les autres outils doivent par conséquent être alignés au bon moment.

Chapter 11

Wordpress

Nous utilisons Wordpress principalement pour gérer nos contenus **H5P**. Cependant, étant basé sur un serveur LAMP (Linux-Apache-MySQL-PHP), il permet également d'implémenter des fonctions basées sur des scripts côté serveur qui seraient difficiles, voire impossibles avec un site statique tel que l'offre {bookdown}, par exemple. Ainsi, nous pouvons aussi y intégrer un enregistrement de l'utilisateur via GitHub (protocole OAuth2). Ceci est bien pratique pour faire le lien entre le compte UMONS/Moodle et le compte Github de l'étudiant.



Nous avons choisi de privilégier l'authentification GitHub à celle de l'UMONS pour les raisons suivantes :

- Certains étudiants arrivent au cours sans être encore complètement inscrits, avec leur propre PC portable. Ils n'ont pas encore de login UMONS et ne pourraient rien faire si on dépendait exclusivement de ce login pour travailler. Avec Github, ils créent leur compte librement et sont immédiatement opérationnels.
- Nos outils sont prêts et utilisables en dehors du contexte de l'UMONS également (si nécessaire, par exemple, pour une formation dans le cadre d'un workshop international).
- Comme GitHub et GitHub Classroom occupent une place importante dans les outils d'apprentissage de nos cours, les étudiants doivent de toute façon créer et utiliser un compte GitHub. L'authentification via GitHub permet d'être certain que ce compte est créé et fonctionnel et que l'étudiant fournit effectivement son propre login.

Toutes ces raisons sont suffisantes pour que l'authentification GitHub soit considérées comme l'élément clé dans le contexte de nos cours de Science des Données. Ainsi pour éviter de multiplier les logins, l'identifiant Wordpress de l'étudiant sera toujours identique à son login GitHub avec cette approche.

En 2018-2019 et 2019-2020, nous avons travaillé uniquement avec un serveur statique {bookdown} hébergé sur GitHub. Si le résultat est correct, nous avons noté plusieurs problèmes :

- Pas d'enregistrement des étudiants (suivi de progression impossible),

- Pas de H5P. Nous pourrions utiliser <https://h5p.org>, mais cette plateforme n'est prévue que pour des tests, et ne permet pas de récupérer les activités des étudiants,
- Pas de service d'applis Shiny et de learnrs. Nous exécutons ces applis en local dans la SciViews Box via les packages R BioDataScience et BioDataScience2. Cela permet d'utiliser les learnrs, mais sans être vraiment très pratique pour les étudiants. Les learnrs qui y sont intégrés se connectent à la base de données MongoDB via une petite appli Shiny intégrée qui demande de rentrer manuellement dans des zones de texte son login GiHhub et son email pour s'identifier. Toutefois, comme aucun contrôle n'est fait sur ce que les étudiants entrent dans ces champs, nous avons eu beaucoup d'enregistrements fantaisistes qu'il a fallu régler grâce à une table de corrections appliquée aux données de la base.

Au final, même si l'approche bookdown statique sur Github + learnr/Shiny en local dans la SciViews Box permet d'obtenir le minimum de fonctions requises pour le suivi de la progression d'apprentissage des étudiants, ce n'est pas pratique à l'usage et nous avons dû dépanner souvent des étudiants... sans compter leur stress de savoir si leur activité est bien enregistrée dans la base de données ou non (dans nos cours à l'UMONS, deux points sur vingt dans la note totale sont réservés à l'exécution de ces learnrs, d'où l'importance de bien enregistrer qui fait quoi).

11.1 Authentification des étudiants

En 2020-2021, nous avons décidé de mettre en œuvre un serveur Wordpress pour fournir le contenu H5P qui nous manque et dont l'utilisation est enregistrée pour chaque étudiant afin de pouvoir suivre sa progression (protocole xAPI). Ce serveur Wordpress nous donne aussi l'occasion de mettre en œuvre une authentification des étudiants plus fiable. Notre objectif est double au niveau de l'authentification de nos étudiants. D'une part, nous voulons obtenir de manière fiable à la fois leurs données UMONS (via Moodle) et leur login GitHub, ce qui implique que ces données doivent être soit fournies automatiquement, soit vérifiées durant l'authentification. D'autre part, nous voulons pouvoir exécuter des applications Shiny en mode sécurisé pour, par exemple, leur offrir un rapport de progression calculé en direct. La nouvelle authentification se fait comme suit :

1. Nous intégrons nos outils dans Moodle via une URL qui transmet des informations relatives à l'utilisateur authentifié UMONS ("query string", c'est-à-dire la partie "?..." de l'URL). On obtient de cette façon, entre autres, son numéro de matricule, son adresse email UMONS, son nom et son prénom, le cours qu'il suit, ... Les détails sont dans la section correspondante (voir 12) et ne seront donc pas discutés ici.
2. Par ailleurs, nous utilisons abondamment GitHub pour les exercices des étudiants. Leur login y est différent, et il est aussi possible qu'ils y aient indiqué une adresse email différente (même si nous leurs demandons de mettre la même, ce qui est d'ailleurs indispensable pour l'identification correcte de l'auteur des commits). Dans Wordpress, nous pouvons intégrer une authentification sur base de leur compte GitHub via le protocole OAuth2 au travers d'une application qui fait le pont entre Wordpress et GitHub, nommé **BioDataScience**.

C'est doublement intéressant car nous obtenons ainsi leur login GitHub à tous les coups, et aussi, nous sommes certains que leur compte GitHub est créé et actif dès le moment où ils sont enregistrés dans Wordpress. Que des avantages, donc. De plus, dans

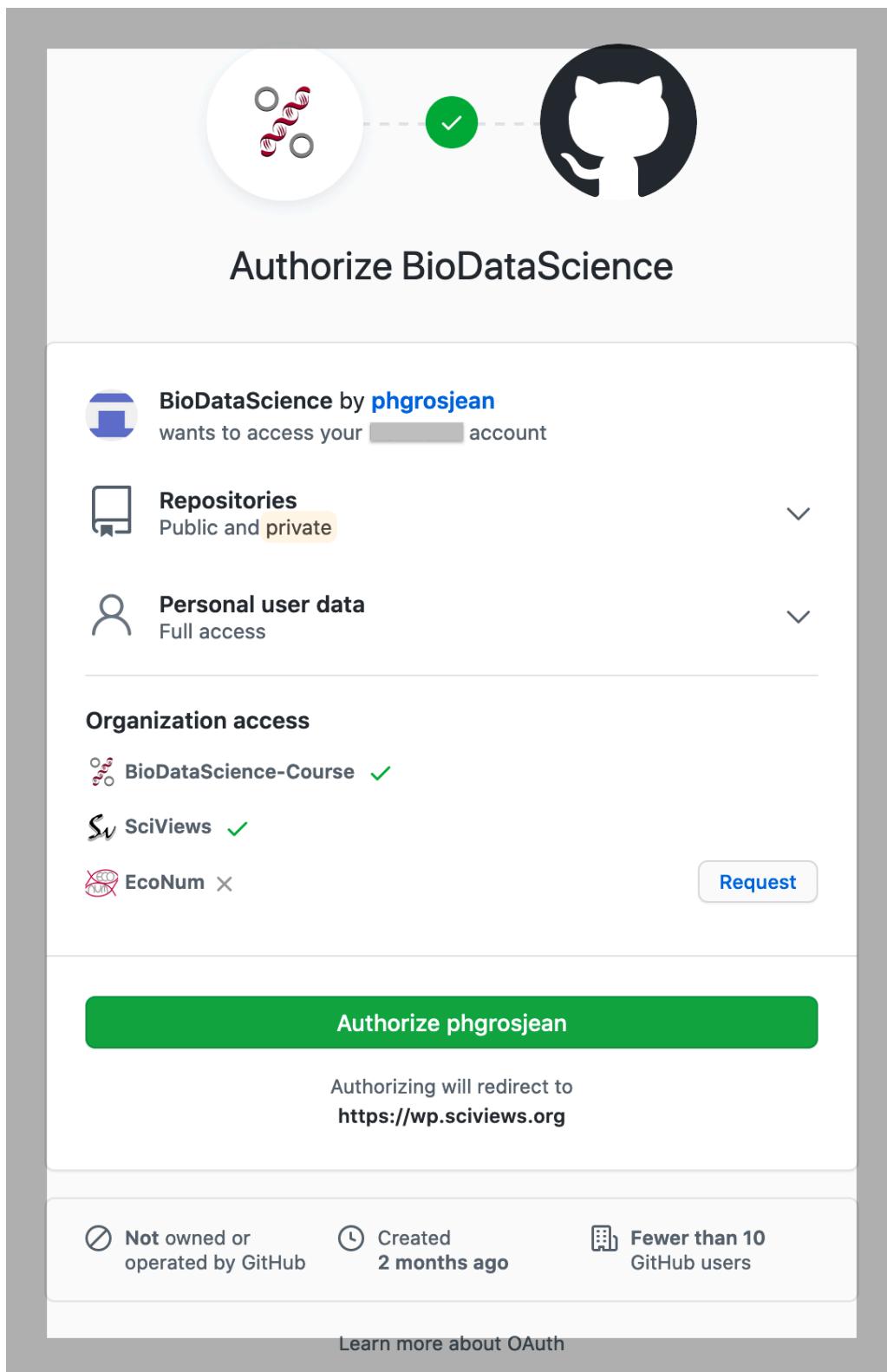


Figure 11.1: Écran de login via GitHub la première fois que l'étudiant se connecte à GitHub via Wordpress au travers de l'application BioDataScience.

l'implémentation choisie, la connexion via GitHub montre clairement à l'étudiant s'il est connecté ou non.

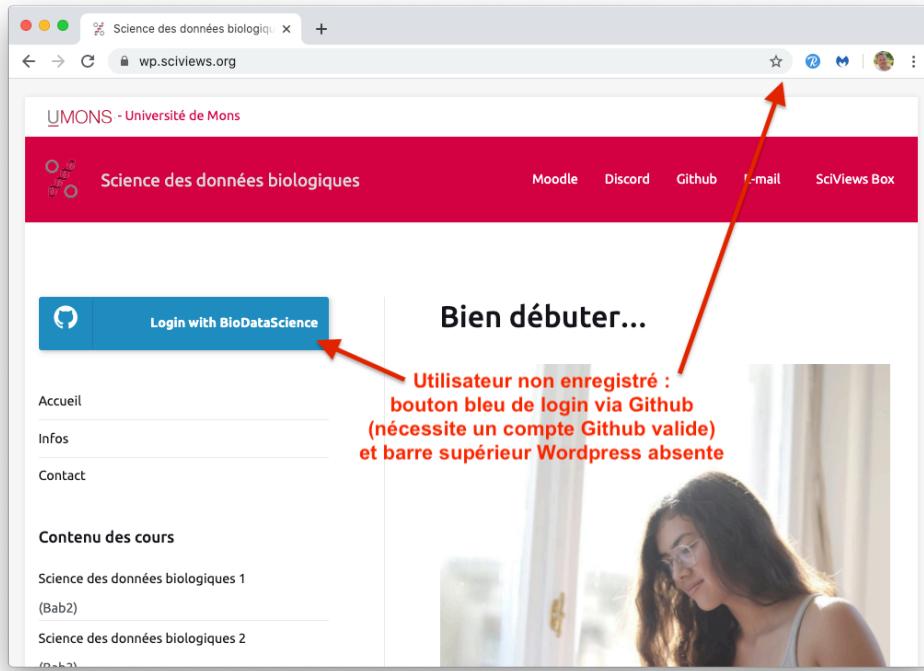


Figure 11.2: Page principale du site lorsque l'utilisateur n'est pas connecté (au travers d'un login BioDataScience dans Github).

Afin d'éviter que n'importe qui ayant un compte GitHub ne s'enregistre dans Wordpress, nos étudiants doivent s'enregistrer au préalable (TODO: formulaire d'enregistrement à ajouter dans la page “bien débuter...”). Ce formulaire crée un tableau CSV qui rassemble les données issues de Moodle et le login GitHub que l'étudiant entre après avoir créé son compte. Ce tableau CSV doit être vérifié par l'enseignant et ensuite, il ajoute les comptes à Wordpress via le plugin **Import users from CSV** qui est installé dans Wordpress. Une fois enregistré, l'étudiant reçoit un mail et la page “bien débuter...” l'indique explicitement.

Dès que l'étudiant s'est enregistré, les données utiles venant de Github sont rentrées dans la table des utilisateurs Wordpress. Nous pouvons donc combiner cette information avec les données provenant de Moodle pour établir un profil complet de l'étudiant dans notre site Wordpress automatiquement. A noter que Wordpress tente de récupérer aussi d'autres informations venant de GitHub, dont l'email. Cependant, ce dernier peut être caché dans le compte GitHub (c'est d'ailleurs le cas par défaut), et n'est donc pas accessible. La plupart du temps, nous n'avons donc que l'email UMONS à disposition. **Nous avons donc décidé d'utiliser l'email Wordpress, l'email UMONS, ou le login Github de manière interchangeable selon le contexte comme identifiant**, mais seule la présence du login est vérifiée pour activer l'enregistrement de l'activité dans H5P/learnr/Shiny. Comme l'email git est mis à celui de l'UMONS lorsque RStudio Serveur est accédé via Moodle, l'étudiant se doit d'indiquer le *même* email dans son compte GitHub, sans quoi, ses commits à partir de la SciViews Box ne pourront pas lui être attribués dans GitHub. *C'est de sa responsabilité et actuellement, nous n'avons*

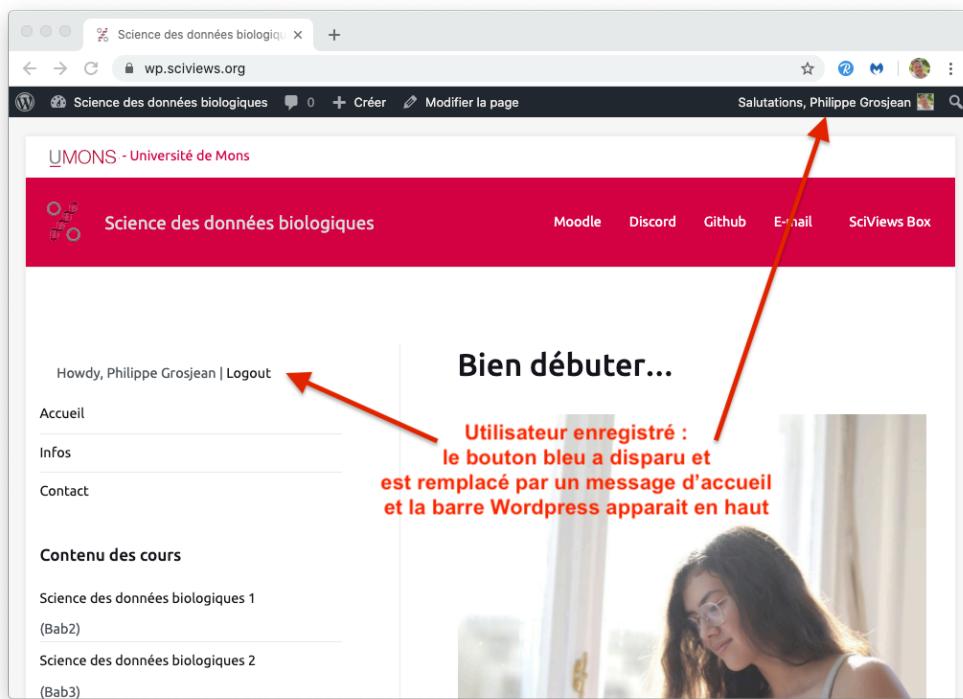


Figure 11.3: La même page principale du site lorsque l'utilisateur est enregistré et connecté via Github.

aucune visibilité sur l'email utilisé sous GitHub par l'étudiant dans l'implémentation mise en place¹.



Le changement automatique d'email git dans la SciViews Box (à partir de la version 2020) ne se fait *que* si RStudio est accédé depuis le lien dans Moodle UMONS. Ainsi, un externe peut très bien utiliser n'importe quelle autre adresse mail sans problèmes.

L'adresse mail utilisée est celle renseignée par Moodle, soit `prénom.nom@[student.]umons.ac.be`. Cependant, dans de rares cas, l'adresse UMONS utilise des caractères qui posent problèmes ailleurs (GitHub en particulier), comme une apostrophe dans le nom. Dans ce cas, il faut basculer sur l'adresse alternative `matricule@umons.ac.be`.

TODO: Utiliser un autre plugin OAuth2 et implémenter l'ajout de l'email lorsqu'un nouveau compte est créé depuis Github.

11.2 Installation Wordpress

Notre nom de domaine `sciviews.org` est géré par <https://gandi.net>. Ce fournisseur propose aussi un hébergement web Wordpress dans un serveur LAMP préconfiguré. Nous utilisons cette option. Le serveur utilise 1 cpu et 256Mo de RAM avec 20Go d'espace de stockage, et un quota de 300.000 pages vues par mois et 60Go par mois. Il peut être augmenté à tous niveaux à n'importe quel moment (plus de cpus ou de ram, plus de débit, et espace disque jusqu'à 1To). Cette configuration se comporte très bien en charge avec nos étudiants (> 170 par an et 75 max par classe). Nous pourrons toujours passer vers un plan supérieur si des lenteurs sont constatées. Le certificat SSL pour avoir une adresse sécurisée en `https://` est inclus dans le pack. Nous ne migrerons pas vers une instance gérée par le Wordpress de l'UMONS car le sous-traitant impose des contraintes qui ne conviennent pas à notre usage (restriction sur les plugins et pas d'utilisation d'iframe). De plus, la gestion se fait via le Département d'informatique qui dialogue avec le sous-traitant -pas d'échange direct avec le sous-traitant-, ce qui introduit un temps de latence insupportable de parfois plusieurs semaines avant d'obtenir une modification) !

Le serveur est installé initialement avec **PHP 7.3** et une base de données **MySQL 5.7** ne permettant que des requêtes locales (impossible d'accès par l'extérieur par décision du fournisseur). En janvier 2022, PHP est passé à la version 7.4. **Wordpress** est installé initialement à la version **5.4.2**. Il a évalué progressivement vers la version **5.8.2** en fr_BE en janvier 2022 qui est la dernière version en date et qui recommande PHP 7.4 minimum et MySQL 5.7. Donc, on est bon. Le site utilise 4Go sur les 20Go disponibles.

Le site est servi à l'adresse <https://wp.sciviews.org>. Les enseignants y sont enregistrés comme éditeurs. En outre, un compte administrateur est aussi créé. Le serveur est accessible par SSH (après activation provisoire de la console dans Gandi), sFTP et git/SSH. Toutes les informations nécessaires à la gestion de l'hébergement sont transmises aux intéressés par voie sécurisée.

¹Pour les enseignants, lors des premiers commits réalisés par vos étudiants, veuillez vérifier s'ils sont clairement identifiables. Sinon, faites corriger l'information dans la SciViews Box (identifiant git et email renseignés en haut à droite dans la boîte de dialogue de configuration de la SciViews Box).

11.2.1 Plugins Wordpress

Nous avons installé les extensions suivantes (les versions correspondent à la configuration de Wordpress au 05/01/2022 et sont toutes à jour à cette date) :

- **H5P 1.15.3.** Il faut installer les libraires séparément et les principales ont été rajoutées dès le départ. Les 5 ou 6 widgets H5P que nous avions déjà créés sont facilement importés en les exportant de leur site original au format .h5p en cliquant sur le bouton **Reuse** en bas à gauche du widget. Ensuite dans Wordpress, on va dans le tableau de bord -> H5P Content, on clique **Add New** puis **Téléverser** et on sélectionne le fichier .h5p. Une sauvegarde de ces fichiers .h5p est également effectué sur p-Cloud dans le dossier **sdd_h5p_backup**. Le plugin est ensuite configuré pour ne pas montrer la bannière H5P en bas des widgets et pour générer tous les événements xAPI possibles (voir 8).
- **SNORDIAN's H5PxAPIkatchu 0.4.11** capture les évènements xAPI générés par H5P et les stocke directement dans la base de données MySQL liée à Wordpress. Il y a une interface simpliste pour visualiser les tables, les filtrer et les exporter au format CSV. Nous avons patché cette version pour y rajouter l'enregistrement des évènements xAPI directement dans notre base de données MongoDB Atlas, et n'utiliser l'enregistrement local original que comme solution d'appoint au cas où l'enregistrement en MongoDB s'avère impossible pour une raison ou une autre (base saturée, indisponible, erreur, ou port 27017 de MongoDB bloqué). **Attention : depuis la version 0.4.7, il faut activer l'option "Embed support"**, sinon, les exercices H5P des pages bookdown ne sont pas enregistrées à partir de certains navigateur Web (testés avec la version 0.4.11 : Safari et Chrome sous MacOS Big Sur, Edge et Chrome sous Windows 10 et Chromium et Firefox sous Ubuntu 20.04LTS).
- **OAuth Single Sign On - SSO (OAuth Client) 6.21.6** configuré pour enregistrer et connecter un utilisateur via son compte Github. Pour permettre ceci une application Github nommée **BioDataScience** a été créée sur Github. La configuration de ce plugin est disponible depuis le tableau de bord de Wordpress -> miniOrange -> OAuth.
- **Import and export users and customers 1.19.1.2** permet d'exporter un tableau contenant tous les utilisateurs enregistrés dans Wordpress, accompagné des métadonnées. Ce tableau CSV est ensuite lu par la fonction **sdd_users()** pour créer les informations utilisateurs accessibles par RStudio Connect et les learnrs/apps Shiny exécutées directement dans la SciViews Box. Sa fonction d'importation des données permet de mettre à jour rapidement des données et métadonnées relatives à un grand nombre d'étudiants en une seule fois à partir d'un tableau CSV.
- **Advanced iFrame Pro 2021.9** permet d'insérer un iframe avec plein d'options : sécurisation, adaptation de la taille, passage de paramètres, réécriture d'URL, etc. Nous utilisons ce plugin pour intégrer les bookdowns dans le site Wordpress, et ainsi conserver la bannière supérieure de notre site tout en affichant le bookdown. La version standard est installée depuis le site plugins Wordpress, et puis nous avons payé \$19 pour obtenir la version PRO qui propose des fonctionnalités en plus dont nous avons besoin pendant 6 mois (renouvellement de 6 mois = \$10). Pour la mise-à-jour, on télécharge la nouvelle version sous forme de fichier zippé depuis Evanto, on supprime l'ancienne version (config conservée), et on installe la nouvelle version

en téléversant le fichier .zip. La configuration est telle qu'elle correspond au code suivant (qui peut être utilisé tel quel sans modifier par ailleurs quoi que ce soit dans les boîtes de dialogue) :

```
[advanced_iframe use_shortcode_attributes_only="true"
src="//www.tinywebgallery.com/blog/advanced-iframe"
width="100%" height="600" add_iframe_url_as_param="same"
show_iframe_loader="true" enable_responsive_iframe="true"
hide_elements=".tg-site-footer-bar,.entry-footer" additional_height="100"
multi_domain_enabled="false" use_post_message="false"]
```

- **My Custom Functions 4.51** permet de rajouter des fonctions PHP au site sans devoir créer de plugin Wordpress pour cela et sans “polluer” les autres fichiers (comme `functions.php` du thème). Les fonctions rajoutées créent des variables dans `localStorage` du browser web sur le client pour l’identification de l’étudiant entre bookdown, H5P et les applis Shiny/learnr, pour enregistrer les évènements H5P xAPI dans MongoDB Atlas via H5PxAPIkatchu, et pour modifier Wordpress de manière cosmétique (pas d’icône Wordpress dans la barre du haut, logo et lien BioDataScience dans la page de login).
- **WP User Avatars 1.4.1** permet de gérer les avatars des utilisateurs. Nous l’employons pour placer les photos de chaque étudiant issues des données UMONS. Pour l’instant, nous devons vérifier chaque utilisateur manuellement et rajouter ces photos à la main... mais au moins, ce plugin permet de le faire.
- **UpdraftPlus 1.16.69** comme système de sauvegarde de l’ensemble (fichiers Wordpress + base MySQL). Le backup se fait en local pour l’instant dans `lamp0/web/vhosts/wp.sciviews.org/htdocs/wp-content/updraft`, mais il est recopié manuellement par sFTP sur notre disque p-Cloud dans le dossier `sdd_site_backup`. **TODO:** il nous faudra à terme une solution de backup automatisé sur p-Cloud ou OneDrive UMONS.
- Quelques autres plugins ne sont pas essentiels, mais aident à gérer le site Wordpress : **Broken Link Checker 1.11.16**, **Health Check & Troubleshooting 1.4.5**, **WP-Optimize 3.2.1**, **WP Rollback 1.7.1** & **WordPress REST API Authentication 1.6.6**.

Tous ces plugins sont distribués sur le site `plugins de Wordpress`, sont à jour et activement maintenus, avec un grand nombre d’utilisateurs. Nous avons évité les plugins plus “confidentiels” pour lesquels nous ne sommes pas certains de la pérennité à plus long terme. Note/ : en janvier 2022, nous avons recherché la compatibilité PHP 8.0 de ces plugins dans la doc, mais sans les tester. Nous avons trouvé des indications de compatibilité pour tous sauf OAuth SSO, Wordpress REST API, WP Rollback, WP User Avatars & Health Check & Troubleshooting. Nous n’avons pas non plus trouvé d’indications de compatibilité pour le thème Zakra utilisé (voir ci-dessous). De plus, il faudra faire une mise à jour des fichiers composer pour l’extension MongoDB pour PHP !

11.2.2 Thème du site

Pour le thème, nous avons choisi Zakra version “free”, à jour, activement maintenu et totalement personnalisable. En jouant sur les options uniquement, nous avons pu façonner

le site de telle manière qu'il reprenne une identité et les couleurs UMONS, tout en restant le plus simple, intuitif et facile d'accès possible. En janvier 2022, la version 2.0.8 est utilisée (dernière en date).

11.3 Configuration serveur et git

Simple hosting de Gandi.net permet d'accéder aux fichiers sur le serveur de trois façons :

- Par sFTP (TODO: détails)
- Par Console SSH (TODO: détails). La console n'est pas active en permanence. Il faut l'activer dans le panneau de configuration du simple hosting sur Gandi.net. Cette activation reste valable pour 2 heures seulement par sécurité.
- Via git. C'est le plus simple pour développer et synchroniser son site. C'est d'ailleurs aussi la seule façon de faire pour ajouter des dépendances à PHP via composer (et nous aurons besoin de rajouter une dépendance par rapport à MongoDB pour enregistrer les évènements H5P xAPI dans notre base de données).

11.3.0.1 Sécurisation du site Wordpress

- Modification du fichier `.htaccess` dans `htdocs`. Afin d'éviter des attaques de type pingback xmlrpc sur le site Wordpress, il faut rajouter la ligne suivante dans le fichier `.htaccess` :

```
RewriteRule ^xmlrpc\.\php$ "http\:\/\/0\.0\.0\.0\/" [R=301,L]
```

- Utilisation des plugins “Health Check & Troubleshooting” et “WP-Optimize” pour diagnostiquer et remédier aux problèmes potentiels. Un seul problème reste impossible à résoudre pour l'instant, il s'agit de “A plugin has prevented updates by disabling `wp_version_check()`.” qui peut avoir de nombreuses origines, et dans le cas de notre site, nous n'avons pas pu détecter la cause. Même en désactivant tous les plugins et le thème, le message reste. Un autre avertissement concerne la conservation de thème inutilisés. En effet, nous conservons les thèmes par défaut de Wordpress “Twenty Twenty” et “Twenty Twenty-One” car il est conseillé de les garder, au cas où Zakra aurait des problèmes. Enfin, nous utilisons aussi “Broken Link Checker”. Cet outil détecte les URL incorrectes dans la pages Wordpress uniquement (pas dans les bookdowns, donc), mais cela aide. Les “issues” étant des liens qui nécessitent une authentification, cet outil les détectent tous les quatre comme incorrects, mais nous pouvons ignorer ces avertissements également. Pour tester les liens dans les bookdowns, nous utilisons un plugin Chrome en plus.
- Nous avons aussi utilisé l'outil de sécurité et d'audit de Chrome dans les outils développeurs pour détecter les failles éventuelles et les erreurs côté client (Javascript, CSS, ...), et nous assurer que le site est sain. Globalement, il n'y a pas de failles majeures, mais certains aspects pourraient être améliorés pour avoir un site plus rapide, par exemple. Nous examinerons ces points plus tard.

11.3.0.2 Utilisation de git

Il faut git et un client SSH installé (pas de problèmes pour le Mac ou Linux). Les infos pour utiliser git sont dans la page de gestion du site wp.sciviews.org sur Gandi.net. Dans un premier temps, l'identification se fait via un mot de passe. Voici comment cloner le dépôt git à partir d'un répertoire racine de votre choix, mais nous aurons absolument besoin que ce répertoire racine soit le même que celui où les dépôts GitHub de BioDataScience-Course sont également placés (sur le Mac, il est facile de s'y placer en tapant cd suivi d'une espace, et puis en allant faire un drag&drop du dossier depuis le Finder vers le terminal, puis d'exécuter cette commande).

```
# Placez-vous dans le bon dossier
#cd <dossier racine des dépôts BioDataScience-Course>
git clone git+ssh://3693753@git.sd3.gpaas.net/wp.sciviews.org.git
```

Accepter (**yes**) pour le message relatif à l'authenticité, puis rentrer le mot de passe d'administration du site. **Vérifier que les fichiers du dépôt git ont bien été rajoutés dans le dossier avant de continuer.** Avant de pouvoir effectuer un push via GitHub Desktop, il faut avoir configuré git correctement et avoir une clé SSH.

```
git config --global user.name <username>
git config --global user.email <user@email.com>
# Générer une clé SSH pour l'authentification
ssh-keygen -t rsa -b 4096 -C '<user@email.com>'
```

Une fois la clé SSH générée, copier le contenu de la clé publique `~/.ssh/id_rsa.pub` d'une part vers "SSH Keys" dans GitHub -> Settings -> SSH and GPG Keys (pour être aussi en ordre vis-à-vis de GitHub), et d'autre part, me la fournir (Philippe Grosjean) via email afin que, dans le panneau principal Simple hosting de Gandi.net (note : dans la section sécurité, cliquer sur "Ajouter une clé SSH". Puis créer une nouvelle clé et y coller le contenu).

Avant de travailler dans un des bookdowns du cours, toujours se mettre à jour en local en faisant un pull dans le dépôt GitHub sdd-umons-xxxx depuis RStudio, ainsi qu'un pull ou un fetch depuis GitHub Desktop sur le dépôt wp.sciviews.org créé ci-dessus Essayez une petite modification anodine dans votre dépôt git en local, puis commit et push sur GitHub. A ce stade, ce sont les *sources* du bookdown qui sont synchronisées. Allez ensuite dans GitHub Desktop au dépôt wp.sciviews.org et faites-y également un commit, et puis un push ou fetch. A ce moment-là, les pages du bookdown sont synchronisées avec le dépôt git de Gandi.net, mais pas encore *publiées* sur le site wp.sciviews.org à ce stade ! Il faut **déployer** ces changements en faisant dans une fenêtre terminal :

```
cd <Répertoire de base qui contient le dossier wp.sciviews.org>
ssh 3693753@git.sd3.gpaas.net 'deploy wp.sciviews.org.git'
```

Le **deploy** ajoute des nouveaux fichiers et remplace les fichiers existants sur le site (à condition que le droit en écriture sur les fichiers soit OK). Par contre, il n'élimine pas de fichiers. On risque donc d'accumuler des fichiers devenus inutiles au fur et à mesure de l'édition des blogdown. Si cela ne devrait pas impacter l'affichage des pages, cela utilise inutilement de l'espace disque sur le serveur. Donc, de temps en temps, il faut procéder à un nettoyage complet avec `clean` suivi d'un redéploiement. **Note : n'utilisez pas encore cette commande ! Laissez-moi (Philippe Grosjean) la tester avant cela.**

```
ssh 3693753@git.sd3.gpaas.net 'clean wp.sciviews.org.git'
```

Attention qu'un fichier .gitignore correct est indispensable ici dans le dépôt wp.sciviews.org. Les fichiers concernés par le .gitignore ne seront pas effacés. Tout le reste oui (faire toujours une sauvegarde complète avant) !

11.3.1 Ajout de mongodb pour PHP

Le serveur sur Gandi.net n'autorise qu'une seule manière de rajouter des dépendances pour PHP : via git + ssh en utilisant `composer`. Et il faut créer les fichiers `composer.json` et `composer.lock` sur une machine locale avant de pouvoir faire un push sur le serveur. Cela signifie d'avoir PHP et la dépendance mongodb pour PHP installés localement. Ceci n'est pas facile, et il faut les mêmes versions que sur le serveur. Comme le PHP sous MacOS Catalina a la version 7.3 de PHP, j'ai d'abord essayé sur le Mac... sans succès car la puce de sécurité T2 du Mac empêche des modifications indispensables dans les dossiers `\usr\...` pour pouvoir compiler mongodb pour PHP. J'ai essayé aussi avec la SciViews Box 2019 (mais version PHP 7.2 trop ancienne) et une VirtualBox Ubuntu Server 20.04, mais version PHP 7.4 trop élevée. Gasp ! Finalement, j'ai pu le faire avec une VirtualBox Xubuntu 19.10. La procédure est détaillée ci-dessous pour mémoire (mais comme elle n'est nécessaire qu'une seule fois, il ne faut plus s'en préoccuper). Plus loin, je mets aussi la procédure utilisée sur un Xubuntu 20.04LTS pour une mise à jour de l'extension...

- Noter la version X.Y de PHP utilisée sur le serveur (ici 7.3). Il faudra la même en local.
- Aller dans MongoDB Atlas et regarder la version de MongoDB déployée (ici 4.2). Il faudra une version de l'extension mongodb qui supporte à la fois la version de PHP et celle de MongoDB qui gère notre base de données.
- Rechercher la version compatible dans le tableau relatif à PHP ici. Nous pouvons voir qu'il nous faut l'extension 1.6+ ou 1.7+, avec la librairie 1.5 ou 1.6, respectivement, et par ailleurs cette version de l'extension est bien compatible avec la version 7.3 de PHP.
- Installer un (X)Ubuntu d'un version telle qu'il propose la bonne version de PHP. La 18.04LTS utilise PHP 7.2, pas bon. La 20.04LTS utilise la version 7.4 de PHP. Au final, une 19.10 convient, et c'est la version la plus récente avant qu'Ubuntu ne passe à PHP 7.4. Une machine virtuelle préinstallée peut être téléchargée depuis Osboxes. La façon de l'installer dans VirtualBox est expliquée sur le site. Le disque virtuel initial de la machine occupe 5,89Go.
- Configurer la machine virtuelle dans VirtualBox pour avoir un dossier partagé, par exemple 'Desktop' partagé en 'MacDesktop', avec connexion automatique.
- Lancer la machine virtuelle. Login : "oxboxes", mot de passe : "oxboxes.org", mais attention, en clavier configuré QWERTY. Avec un AZERTY, il faut entrer en fait "oxboxes:org" comme mot de passe.
- Aller dans le panneau de configuration et changer le clavier (belge, ou français (Macintosh) par exemple).
- Lancer un terminal et exécuter les lignes suivantes :

```
# Rendre le dossier partagé VirtualBox accessible pour l'utilisateur oxboxes
sudo usermod -aG vboxsf osboxes
# Mettre la machine à jour
sudo apt update
sudo apt upgrade
sudo apt autoremove
sudo apt autoclean
sudo apt clean
# Redémarrer
sudo reboot
```

- Il se peut que la machine virtuelle rebascule en QWERTY. Vérifier et corriger à nouveau. Ensuite, installer PHP et ses dépendances nécessaires :

```
sudo apt install php7.3-cli php7.3-dev php-pear php-mongodb composer
```

Ensuite, nous devrions pouvoir créer les fichiers nécessaires via la commande `composer require mongodb/mongodb`, mais cela ne fonctionne pas lorsque `php-mongodb` est installé via le package Ubuntu. Nous devons le compiler nous-même avec `pecl` sur la machine² :

```
# Créer un répertoire dans le dossier partagé et s'y placer
mkdir /media/sf_MacDesktop/php-mongodb
cd /media/sf_MacDesktop/php-mongodb
# Compiler mongodb depuis les sources
sudo pecl install mongodb
# Créer composer.json et composer.lock (entre autres)
composer require mongodb/mongodb
```

Pour Xubuntu 20.04LTS et mongodb dernière version pour PHP 7.4, je fais :

```
sudo apt install php7.4-cli php7.4-dev php-pear php-mongodb composer
mkdir ~/Desktop/php-mongodb
cd ~/Desktop/php-mongodb
sudo pecl install mongodb # Fails because it is not indicated in php.ini
# But the compilation of the .so file did succeed
composer require mongodb/mongodb # Warnings about deprecated composer 1 ignored
```

- Copier `composer.json` et `composer.lock` ainsi obtenus à la racine du dépôt git, faire un commit, un push et un deploy (`ssh 3693753@git.sd3.gpaas.net 'deploy wp.sciviews.org.git'`). Cette fois-ci mongodb PHP 1.6 semble s'être bien installé sur le serveur lors du deploy, mais pour l'upgrade à 1.11, il m'indique que je n'ai pas `mongodb-ext` 1.12.0, mais 1.10.0. Ceci peut aussi être vu à la console avec :

```
#php --ini
php -m # mongodb must be listed
php --re mongodb | head -1
```

² Je suis arrivé à créer `composer.lock` via la commande `composer update` en ayant été écrire à la main un `composer.json` qui contient les lignes minimales renseignées sur la page d'aide de Gandi.net, mais j'aboutis à une version trop ancienne du driver, et de toutes façons, c'est un bricolage qui ne me plaît pas !

Cela signifie que la version max. de mongodb lib qui peut être installée avec mon PHP 7.4 est la 1.9 si je ne veux pas mettre à jour également mongodb-ext. Donc :

```
sudo pecl uninstall mongodb
sudo pecl install mongodb-1.10.0
composer require mongodb/mongodb ^1.9
```

Maintenant, j'ai les bons `composer.json` et `composer.lock` pour une version reconnue par MongoDB 5.0 (avec une limitation en lecture, mais ici, on n'accède que en écriture). Faire commit, push et deploy et vérifier que tout se passe bien.

- Faire un test dans Wordpress. Aller jusqu'à une contenu H5P quelconque sur le site et interagir avec lui. Aller ensuite voir si les données correspondantes ont bien été enregistrées dans notre MongoDB ATLAS, dans la collection `sdd.h5p`. Si pas, vérifier si ces mêmes données sont enregistrées dans `H5PxAPIkatchu` dans le tableau de bord Wordpress. Si c'est le cas, c'est un problème d'accès à la base de données MongoDB ATLAS, sinon, c'est un problème plus profond : est-ce que l'enregistrement xAPI est bien activé, et est-ce que les extensions bookdown nécessaires ont bien été injectées grâce à une configuration correcte de learnitdown ? L'option de débogage de `H5PxAPIkatchu` devrait s'avérer utile ici pour déterminer ce qui ne va pas, voir 8.

11.4 Agencement du site

La page principale <https://wp.sciviews.org> sert essentiellement à configurer le site lorsqu'elle est appelée depuis Moodle (enregistrement des données de l'étudiant dans un `localStorage`) et donne les instructions principales pour bien démarrer avec le contenu pédagogique.

Cette page laisse une large place à gauche pour le login GitHub, quelques liens utiles (acceuil, contact) et liste les différents bookdown disponibles, un par cours.

Chaque bookdown est dans une sous section du site, et reprend un numéro de version correspondant au début de l'année académique : `/sdd-umons-2020/`, `/sdd-umons2-2020/`, etc. pour l'année académique 2020-2021, par exemple. Une fois la page principale activée depuis le lien Moodle **Cours**, l'étudiant peut aller vers le bookdown qui correspond à son cours et tout devrait fonctionner correctement. La page principale de chaque bookdown effectue de son côté les vérifications nécessaires, affiche un court compte-rendu, et permet d'effacer les données personnelles (compatibilité RGPD) si souhaité, voir 5.

Le site est complété d'une bannière supérieure dont l'aspect rappelle le site UMONS et qui propose des liens vers différents items importants dans le cadre du cours pour obtenir de l'aide (Moodle, Issues, Email, ...), et donne aussi accès à l'explication pour installer le logiciel (SciViews Box). Et c'est tout. Le but est de faire ici simple et efficace...

11.4.1 Tableau de bord

Dès qu'un utilisateur s'enregistre sur le site, une barre supérieure noire apparaît avec des options supplémentaires en fonction de son rôle. Un étudiant a évidemment un nombre d'options limitées à l'édition de son profil et la consultation de sa progression H5P. Un

éditeur (prof) peut aussi créer et modifier des pages. Enfin, un administrateur a accès à l'ensemble des fonctions pour personnaliser le site.

Pour tous les utilisateurs, un **tableau de bord** est accessible à partir de cette barre noire supérieure. Une fois le compte créé dans Wordpress, un étudiant a la possibilité de s'y connecter via GitHub et de modifier ses informations dans le tableau de bord, dont son adresse email. S'il le fait, cette adresse n'est alors plus la même que celle de l'UMONS. C'est pourquoi, nous conservons ces deux informations séparément : `email` pour l'email du site Wordpress et `iemail` pour l'email institutionnel tel que renseigné à partir de Moodle. C'est toutefois l'adresse email enregistrée dans Wordpress qui est employée pour enregistrer les activités H5P. Pour les learnr et les app Shiny, c'est le login qui fait référence et qui doit être renseigné, sinon aucun enregistrement n'est réalisé (visiteur externe). Lorsque l'activité learnr ou app Shiny ne se fait, cela est indiqué par un message explicite pour éviter qu'un étudiant oublie de se logger, et puis que son travail soit perdu. Par contre, H5xAPIkatchu enregistre l'activité H5P sous un UUID créé à la volée. Nous pouvons toujours filtrer et éliminer ces données-là plus tard.

Par ailleurs, les évènements principaux H5P (début, fin, durée et résultat à la soumission des exercices) sont également enregistrés dans Wordpress directement. Cette information peut paraître redondante avec l'enregistrement H5PxAPIkatchu, mais elle est utile car l'étudiant ou l'enseignant y a accès directement dans son tableau de bord. Il peut y lire les détails de sa progression dans l'ensemble des exercices H5P de tous les cours, s'il le souhaite (voir capture d'écran). Reste à déterminer l'occupation disque de cette fonction afin de décider si nous la conservons activée sur le long terme !

Content	Score	Maximum Score	Opened	Finished	Time spent
drag-words	3	3	26 May 2020 17 h 31 min	26 May 2020 17 h 31 min	0:15
multiple-choice	1	1	26 May 2020 17 h 31 min	26 May 2020 17 h 31 min	0:04

Figure 11.4: Suivi de la progression H5P dans le tableau de bord accessible aux étudiants.

Comme nous pouvons le voir, le tableau de bord de l'étudiant est réduit au minimum. En plus de la partie "H5P Content" où il peut voir et filtrer la liste des applis H5P et voir ses propres résultats H5P, il a accès à son profil où il peut modifier des données utilisées par nos outils. Il me semble que la maîtrise et la visibilité des données enregistrées le concernant font partie des contraintes RGPD pour lesquelles Wordpress assure la compatibilité.

Les enseignants sont enregistrés comme éditeurs sur le site. Dès lors, ils peuvent aussi

modifier le contenu du site, et surtout, modifier et créer du contenu H5P avec des éditeurs conviviaux. Ainsi, l'ensemble des outils nécessaires pour l'édition du contenu sont regroupés dans ce tableau de bord... à l'exception des bookdowns et des applis Shiny/learnr. Ces dernières sont mise à disposition via un serveur RStudio Connect, totalement différent donc. Par contre, les bookdown doivent, eux, être intégrés dans le site. La façon de réaliser cette intégration est expliquée dans la section consacrée à bookdown (5).

11.4.2 Gestion des événements xAPI

Pour les enseignants, les événements xAPI générés par H5P (voir 8) qui n'ont pas pu être enregistrés dans la base de données MongoDB ATLAS dédiée pour une raison ou une autre sont accessibles à partir du tableau de bord de Wordpress dans la section H5PxAPIkatchu. De là, ils peuvent être exportés au format CSV, et réinjectés dans MongoDB à l'aide d'un peu de code R. **TODO:** H5PxAPIkatchu ne fournit **pas** la date et l'heure de chaque évènement via ce tableau, mais l'enregistre dans la base de données MySQL avec une précision de la seconde. Il faudra peut-être un accès directe à cette base de données pour récupérer l'information complète dans ce cas. Dans MongoDB ATLAS, cette information est enregistrée, par contre, avec une précision de la milliseconde.

Actor Id	Actor Name	Actor Group Members	Verb Id	Verb Display	Object Id	Object Def. Name	Object Des
account: 1fa4415-2568-4671-b7d-9558026d1f4b (https://wp.sciviews.org)			http://adlnet.gov/expapi/verbs/attempted attempted	attempted attempted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=7	multiple-choice	
account: 1fa4415-2568-4671-b7d-9558026d1f4b (https://wp.sciviews.org)			http://adlnet.gov/expapi/verbs/attempted attempted	attempted attempted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=7	multiple-choice	
account: 1fa4415-2568-4671-b7d-9558026d1f4b (https://wp.sciviews.org)			http://adlnet.gov/expapi/verbs/attempted attempted	attempted attempted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=8	R markdown	
account: 1fa4415-2568-4671-b7d-9558026d1f4b (https://wp.sciviews.org)			http://adlnet.gov/expapi/verbs/attempted attempted	attempted attempted	https://wp.sciviews.org/wp-admin/admin-ajax.php?action=h5p_embed&id=7	multiple-choice	

Figure 11.5: Événements xAPI dans H5PxAPIkatchu. En bas du tableau, on peut filtrer les données, et le bouton **Download** permet de récupérer le tableau filtré au format CSV. Le lien **Delete** permet de tout effacer et n'est disponible que pour un administrateur du site.

11.5 Variables contextuelles

Pour le bon fonctionnement de nos outils et l'enregistrement de l'activité des étudiants, nous avons besoin de définir un certain nombre de variables reprises dans le tableau ci-dessous. Toutes ces variables sont obtenues directement à la source³. La colonne origine indique d'où nous obtenons ces informations (pour Moodle, il faut que l'étudiant ait lancé <https://wp.sciviews.org> depuis son compte Moodle et à partir de la page de cours correspondante).

Variable	Contenu	Origine
login	Le login GitHub de l'étudiant	WP via GitHub
email	Le mail renseigné dans Wordpress	WP via GitHub si public
displayname	Nom utilisateur Wordpress(/Github)	WP
firstname	Prénom comme indiqué dans Wordpress	WP
lastname	Nom de famille comme indiqué dans Wordpress	WP
iemail	Le mail UMONS (institutionnel) de l'étudiant	Moddle <code>iemail =</code> Adresse de courriel
iid	Numéro de matricule de l'étudiant	Moodle <code>iid =</code> Nom d'utilisateur
ifirstname	Prénom comme indiqué dans Moodle	Moodle <code>ifirstname =</code> Prénom
ilastname	Nom de famille comme indiqué dans Moodle	Moodle <code>ilastname =</code> Nom
institution	Institution (ici UMONS)	Moodle <code>institution =</code> Institution
icourse	Nom de code Moddle du cours (ex.: S-BIOG-006)	Moodle <code>icourse =</code> N° d'identification du cours
iurl	URL du serveur institutionnel (Moodle)	Moodle <code>iurl =</code> URL du serveur
iref	Identifiant unique dans Moodle	Moodle <code>iref =</code> Numéro d'identification
ictitle	Titre du cours (ex. Science des Données I: visualisation)	Moodle <code>ictitle =</code>
icflag	Indicateur pour les étudiants suivants plusieurs cours simultanément	WP (éditeur des utilisateurs en mode admin)

³En 2018-2019 et 2019-2020, nous utilisions le login GitHub et l'adresse email UMONS pour identifier les événements learnr, mais comme les étudiants devaient rentrer ces infos manuellement, nous avons eu beaucoup de soucis avec des encodages erronés... Donc maintenant, nous récupérons ces infos de manière automatique à la source autant que possible.

Variable	Contenu	Origine
user_login2	Second login GitHub pour le même utilisateur	WP (éditeur des utilisateurs en mode admin)

A noter que nous croisons différentes sources (Moodle, GitHub, Wordpress), et les infos peuvent différer entre ces sources. Considérant un étudiant qui a “installé” ses outils correctement, à savoir :

- Son compte est créé au préalable dans Wordpress via un enregistrement approprié (voir au début de la section),
- Il a créé un compte GitHub et y a renseigné son adresse mail UMONS,
- Il a au moins une fois lancé la page <https://wp.sciviews.org> depuis le lien Moodle “Cours” dans la section “Ressources en ligne” de son cours,
- Dans la page qui s’ouvre, il s’est connecté dans Wordpress à partir du bouton de login via GitHub/BioDataScience.

A ce moment-là, tout sera en place pour gérer correctement son identité dans les différents matériels pédagogiques (bookdown, H5P, learnrs, apps Shiny, etc.). Le seul point sur lequel il faudra rester vigilant est l’identifiant de l’étudiant pour ses commits, pour lequel nous n’avons pas de contrôle automatique pour l’instant. Il ne lui restera plus alors qu’à installer la SciViews Box pour être pleinement opérationnel. Cela semble être un nombre important d’opérations à réaliser pour arriver à cela, mais l’étudiant est guidé de manière naturelle : il va dans Moodle, se connecte et sélectionne le cours dans la liste (il a déjà l’habitude). A partir de là, il clique naturellement sur “Cours”, d'où la page qui l’enregistre et lui propose de créer un compte GitHub. Une fois que c'est fait, il peut cliquer sur le gros bouton bleu “GitHub/BioDataScience” pour se connecter dans le site “Science des Données Biologiques”, et il sélectionne le manuel correspondant à son cours dans la liste. A ce stade, c'est fait. Il est enregistré⁴ !



Faut-il *forcer* le login de l’étudiant dans Wordpress ? Autrement dit, il n’aurait accès aux cours que s’il est connecté. Il existe des plugins Wordpress qui permettent de faire cela, et aussi d’exclure les pages qui peuvent être vues par tous. Un de ces plugins est **Force login**. Nous verrons à l’usage...

Toutes les informations de son compte sont accessibles à l’étudiant dans Wordpress, et il peut librement les modifier dans le tableau de bord, sauf son **login**. Donc, cela signifie que l’information cruciale pour identifier un étudiant est **login** qui est par ailleurs toujours le même que le login GitHub, et **iemail**, son adresse email de l’UMONS fournie par Moodle. Nous avons éventuellement besoin aussi de **email**, son email Wordpress sous lequel les activités H5P sont enregistrées.

Si nous continuons à utiliser Discord, ce serait bien d’avoir le login Discord de l’étudiant aussi. Mais ici, il a le droit d’y poser des questions de manière anonymisée (c’est notre souhait). Néanmoins, il serait utile d’avoir un champ où nous pouvons éventuellement rentrer cette info à la main si souhaité dans la liste des utilisateurs enregistrés dans

⁴Il est conseillé aux enseignants de tester la procédure auparavant pour vérifier qu’ils en maîtrisent bien la logique.

Wordpress (TODO: vérifier si c'est possible). Une autre colonne `comment` serait aussi utile. Nous pourrions y indiquer par exemple si un étudiant a des difficultés particulières, ou la date de son arrivée pour les étudiants étrangers en passerelle, par exemple.

11.6 Redirection d'URL d'iframe bookdown

Les contenus de type {bookdown} forment l'essentiel de l'information contenue dans le site. Cependant, ces pages sont générées indépendamment des pages Wordpress. Elles sont servies à partir du même domaine, mais pas réellement intégrées dans le site Wordpress. Pourtant, visuellement, elles apparaissent comme faisant partie intégrante du site. Leur intégration se fait via un iframe géré par le plugin “Advanced Iframe Pro” qui fait des merveilles pour cette intégration.

Une conséquence visible de cette organisation est l'URL qui présente de manière explicite le fait que la page bookdown est contenue dans un iframe. Par exemple pour <https://wp.sciviews.org/sdd-umons-2020/nuage-de-points.html> en présentation directe hors iframe, la même page est intégrée au site Wordpress à l'intérieur d'un iframe à l'adresse URL suivante : <https://wp.sciviews.org/sdd-umons/?iframe=wp.sciviews.org/sdd-umons-2020/nuage-de-points.html>. Jusqu'ici tout va bien. L'adresse est un peu longue, mais ce n'est pas tellement grave en pratique.

Tout fonctionne bien, à part une petite subtilité concernant les **ancre**s. Une ancre est un signet placé dans une page et qui permet de l'afficher à partir de cet endroit au lieu d'afficher juste le haut de la page. Par exemple, la page `nuage-de-points.html` ci-dessus peut être affichée à partir du point “2.1.1 Echelles de graphiques” correspondant à l'ancre `echelles-de-graphiques` à l'aide de l'URL suivante : <https://wp.sciviews.org/sdd-umons-2020/nuage-de-points.html#echelles-de-graphiques>. Mais si nous rajoutons `#echelles-de-graphiques` à la fin de l'URL avec iframe, il ne se passe rien. La page s'affiche au début, et non pas au point 2.1.1 comme demandé. En fait l'ancre est interprétée pour la page de base Wordpress, pas pour la page contenue dans l'iframe. Pour corriger cela, il faut **URLencoder** l'ancre, qui consiste à remplacer le `#` par son code qui est `%23`. Ainsi, si on écrit <https://wp.sciviews.org/sdd-umons/?iframe=wp.sciviews.org/sdd-umons-2020/nuage-de-points.html%23echelles-de-graphiques>, alors l'ancre fait maintenant partie de l'URL de l'iframe et cela fonctionne comme prévu.

Vous noterez aussi que les ancre ne sont pas rajoutées automatiquement à l'URL lorsque vous cliquez dessus dans une page {bookdown} contenue dans un iframe. **Vous devez donc effectuer le rajout à la main si nécessaire pour enregistrer correctement cette URL.** En effet, “Advanced Iframe Pro” ne gère pas encore correctement les ancre de manière automatique. Une façon simple de découvrir l'identifiant d'une ancre consiste à cliquer bouton droit dessus et à choisir “Ouvrir dans un nouvel onglet” dans le menu contextuel. A ce moment, la redirection du lien cliqué dans l'iframe s'affiche dans une page séparée *hors iframe*, avec son URL complète, y compris son ancre. Copiez cette ancre et collez-là à la fin de l'URL de la page de départ avec iframe après modification du `#` en `%23`, et le tour est joué.

A part ces petits tracas relatifs aux ancre, les pages {bookdown} dans les iframes se comportent très bien par ailleurs et permettent de combiner le meilleur des deux mondes : Wordpress et ses nombreuses fonctions et plugins et {bookdown} pour un contenu riche créé à partir de documents R Markddown.

Chapter 12

Moodle

...TODO: expliquer les URLs spécialisées à ajouter dans les cours pour pouvoir récupérer les informations institutionnelles des étudiants et pour lier notre contenu pédagogique du mieux possible à l'espace Moodle qui y est consacré.

12.1 Gestion des cours

Un espace différent a été créé pour chaque AA (donc, deux pour SDD I et SDD II) par l'administrateur Moodle. Voici la liste des AA, l'intitulé et le code Moodle correspondant (qui ne correspond **pas** au code dans le programme des cours) :

Cours	Code Moodle	Intitulé
SDD I Q1	S-BIOG-006	Science des données I : visualisation
SDD I Q2	(S-BIOG-027)	Science des données I : inférence (pas créé, utilisation de l'espace S-BIOG-006 pour les deux AA)
SDD II Q1	S-BIOG-015	Science des données II : modélisation
SDD II Q2	(S-BIOG-061)	Science des données II : analyse (pas créé, utilisation de l'espace S-BIOG-015 pour les deux AA)
SDD III Q1	S-BIOG-025	Science des données III : exploration et prédiction
SDD IV all	S-BIOG-043	Science des données IV : pratique
SDD V Q1	S-BIOG-077	Science des données V : recherche reproductive
SDD I Charleroi	S-BIOG-921	Bio-informatique et science des données
SDD II Charleroi	S-BIOG-937-	Bio-informatique et science des données II

12.2 Badges

Les badges sont très utiles car ils matérialisent la progression des étudiants. C'est à la fois une façon ludique d'apprendre, et un moyen efficace de pousser un étudiant à travailler.

Les badges sont générés dans Moodle, mais bien sûr, ils sont liés à des activités Moodle. Avec nos applications externes, nous ne pouvons fournir des badges que manuellement, ce qui devient très vite laborieux. **TODO:** trouver un moyen d'automatiser l'octroi de badges. Moodle peut se connecter à <https://badgr.io>, et nous pouvons peut-être travailler via son interface ? A creuser !

Pour créer des badges, nous pouvons utiliser une application en ligne. Il suffit d'indiquer son adresse email, et de concevoir son badge. La même interface peut d'ailleurs être utilisée pour concevoir visuellement des hex-stickers (à faible résolution toutefois puisque c'est des fichiers .png de 200x200 pixels). Voici un exemple de badge que nous pourrions utiliser dans nos cours :

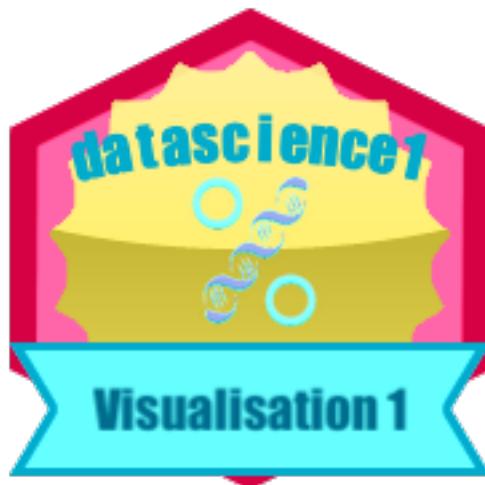


Figure 12.1: Exemple de badge

Les images correspondant à nos différents badges sont sur p-Cloud, dans le sous-dossier `sdd_badges`.

12.3 Notes

TODO: il est possible d'envoyer les notes à Moodle via une URL qui renvoie un fichier XML... Explorer cette piste pour créer et mettre à jour les notes Moodle depuis nos applis...

Chapter 13

Posit Connect

RStudio Connect permet de partager des documents R Markdown, des applications Shiny et des tutoriels Learnrs facilement depuis un serveur centralisé. C'est pratique pour inclure ces items dans le cours en ligne et pour permettre aux étudiant d'effectuer des exercices sans qu'ils n'aient forcément accès à un système complet. RStudio Connect est payant (plusieurs dizaines de milliers de dollars par an), **mais est gratuit pour une utilisation dans un cadre académique. Il faut montrer via un syllabus que l'outil est utilisé dans le cours et que les étudiants apprennent à employer les outils RStudio. La License doit être renouvelée annuellement et n'est valable que pendant la durée du cours (mieux vaut donc avoir des cours qui couvrent toute l'année). À partir de 100 utilisateurs, il est aussi possible d'utiliser l'interface plumber API. Le logiciel doit être installé sur un serveur que l'on configure et administre nous-même.

13.1 Installation de sdd.umons.ac.be

Nous avons opté pour un serveur sdd.umons.ac.be (193.190.194.75) ayant les caractéristiques suivantes :

- CPU Intel Core i7-6900k 3,20GHz à 8 cœurs/16 threads,
- 64Go de mémoire RAM DDR4 2.666Mhz,
- Carte mère Asus X99-A,
- 512Go de disque SSD NVMe Samsung Pro 960 + 2 fois 3To de HDD Toshiba P300 7200rpm SATA III,
- AIO Thermaltake Water 3.0 Performer C,
- Alimentation Coolermaster GM650 650W,
- Carte graphique NVidia GeForce 710 1Go,
- Boitier IPC server 3U-30248,
- Xubuntu 18.04.5LTS,
- MongoDB 4.4.11, MongoDB Tools 100.5.1,
- RStudio Connect 1.8.8.2 + R 4.0.5 (licence 100 utilisateurs qui expire au 2022-05-06),
- Accessible par ssh seulement en interne (ou via le VPN UMONS).

L'installation s'est fait comme suit sur base d'un Xubuntu 16.04 qui a été upgradé à 18.04 :

- Installation de R 3.6.3 (équivalent à svbox2020) :

```
sudo apt install curl
export R_VERSION=3.6.3
curl -O https://cdn.rstudio.com/r/ubuntu-1804/pkgs/r-${R_VERSION}_1_amd64.deb
sudo gdebi r-${R_VERSION}_1_amd64.deb
```

- Installation de R 3.5.3 et 3.4.4, respectivement équivalents à svbox2019 et svbox2018 :

```
curl -O https://cdn.rstudio.com/r/ubuntu-1804/pkgs/r-3.5.3_1_amd64.deb
sudo gdebi r-3.5.3_1_amd64.deb
curl -O https://cdn.rstudio.com/r/ubuntu-1804/pkgs/r-3.4.4_1_amd64.deb
sudo gdebi r-3.4.4_1_amd64.deb

# Test
/opt/R/${R_VERSION}/bin/R --version
```

- Configuration de Java pour R :

```
sudo /opt/R/3.6.3/bin/R CMD javareconf
```

- Installation de RStudio Connect 1.8.2.1-12 :

```
cd ~/Downloads
curl -LO https://rstudio-connect.s3.amazonaws.com/rsc-installer.sh
sudo bash ./rsc-installer.sh 1.8.2.1-12
# Clean up
rm rsc-installer.sh
sudo rm rstudio-connect_*.deb
```

- Installation des dépendances :

```
sudo apt install build-essential libcurl4-gnutls-dev openjdk-7-* libxml2-dev libssl-dev texlive
# Pour les packages R
sudo apt install libgmp10-dev libgs10-dev libnetcdf6 libnetcdf-dev netcdf-bin libdigest-hmac-perl
sudo apt install libgl1-mesa-dev libglu1-mesa-dev libglpk-dev tdsodbc freetds-bin freetds-common
sudo apt install odbc-postgresql libtiff-dev libsndfile1 libsndfile1-dev libtiff-dev tk8.5 tk8.5-dev
```

- Configuration de RStudio Connect/ :

```
sudo mousepad /etc/rstudio-connect/rstudio-connect.gcfg
# URL temporaire du server... https://rstudio-connect-sdd.umons.ac.be
# SenderEmail = sdd@sciviews.org
# Port :3939 (valeur par défaut)
```

- Introduction de la clé d'activation :

```
sudo /opt/rstudio-connect/bin/license-manager deactivate
sudo /opt/rstudio-connect/bin/license-manager activate YAG3-B2H5-MXQB-6HJS-E7WH-4G9Y-JITA
```

- Redémarrage de RStudio Connect :

```
sudo systemctl restart rstudio-connect
```

- Création du premier compte (admin). Naviguer vers <http://en-coris.umons.ac.be:3939>. S'enregistrer et créer le nouveau compte. Configuration également de l'envoi de mail (gandi.net SMTP serveur sur port 587 temporairement).
- Installation de Python :

```
[Python]
Enabled = true
; Python 2.7.12 installed with Ubuntu 16.04
Executable = /usr/bin/python
; Python 3.5.2 installed with Ubuntu 16.04 (and used by svbox2018)
Executable = /usr/bin/python3
```

- Packages Python supplémentaires (pip et pip3 déjà installés) :

```
sudo pip install setuptools
sudo pip install virtualenv
sudo pip3 install setuptools
sudo pip3 install virtualenv
```

- Il faut Python 3.6.7 pour svbox2019 => compilation depuis les sources :

```
# Required dependencies
sudo apt install libffi-dev libgdm-dev libsqlite3-dev libssl-dev zlib1g-dev

# Download and extract source code
cd ~/Downloads
export VERSION=3.6.7 PYTHON_MAJOR=3
wget https://www.python.org/ftp/python/${VERSION}/Python-${VERSION}.tgz
tar -xzf Python-${VERSION}.tgz
cd Python-${VERSION}

# Build Python from source
./configure \
    --prefix=/opt/Python/${VERSION} \
    --enable-shared \
    --enable-ipv6 \
    LDFLAGS=-Wl,-rpath=/opt/Python/${VERSION}/lib,--disable-new-dtags \
    --enable-optimizations
make

# Install Python
sudo make install
```

```
# Clean up
cd ..
sudo rm -rf Python-${VERSION}
rm Python-${VERSION}.tgz

# Check Python install
/opt/Python/${VERSION}/bin/python${PYTHON_MAJOR} --version

# Install pip, setuptools and virtualenv in this Python version
wget https://bootstrap.pypa.io/get-pip.py
sudo /opt/Python/${VERSION}/bin/python${PYTHON_MAJOR} get-pip.py
rm get-pip.py
sudo /opt/Python/${VERSION}/bin/pip install virtualenv
sudo /opt/Python/${VERSION}/bin/pip install setuptools
```

- Ajouter cette version de Python dans la configuration de RStudio Connect (ajouter Executable = /opt/Python/3.6.7/bin/python3 dans la section “[Python]”):

```
sudo mousepad /etc/rstudio-connect/rstudio-connect.gcfg
```

- Redémarrer RStudio Connect :

```
sudo systemctl restart rstudio-connect
```

- Installation du certificat SSL. Copie des fichiers sdd_umons_ac_be.* vers /etc/ssl/private et changement des droits d'accès :

```
sudo chmod o= /etc/ssl/private/sdd_umons_ac_be.cer|crt|csr|key|pfx
```

- Le fichier .cer ne fonctionne pas, mais je peux créer un .pem à partir du .crt au format PKCS#7 :

```
sudo openssl pkcs7 -print_certs -in /etc/ssl/private/sdd_umons_ac_be.crt -out /etc/ssl/private/
sudo chmod o-r /etc/ssl/private/sdd_umons_ac_be.pem
```

- Modification du fichier de configuration de RStudio Connect. Utilisation du port :443, le .key comme clé privée et le .pem comme certificat, et TSL minimum 1.1 :

```
sudo mousepad /etc/rstudio-connect/rstudio-connect.gcfg
```

- Redémarrage du serveur

```
sudo systemctl restart rstudio-connect
```

- Vérification du fichier log

```
sudo cat /var/log/rstudio-connect.log
```

- Synchronisation de /data2/ avec /data1/ :

```
sudo crontab -e
```

Édition du fichier en y ajoutant :

```
# Sync /data1/ with /data2/ every hour:45
45 * * * * /usr/bin/rsync -av --delete /data1/ /data2/ &> /data/datasync.log
```

- Déplacement des données de RStudio Connect vers /data1 :

```
sudo systemctl stop rstudio-connect
sudo mkdir /data1/rstudio-connect
sudo rsync -av --delete /var/lib/rstudio-connect/ /data1/rstudio-connect/
```

Éditer à nouveau le fichier de configuration de RStudio Connect (DataDir = /data1/rstudio-connect) :

```
sudo mousepad /etc/rstudio-connect/rstudio-connect.gcfg
```

Redémarrage de RStudio Connect :

```
sudo systemctl start rstudio-connect
```

- Backup du système en utilisant rear, voir <https://tecmint.com/rear-backup-and-recover-a-linux-system/>

```
sudo apt install rear extlinux
```

- Formatage d'une clé USB :

```
sudo rear format /dev/sdX # In case of error rear format -- --efi /dev/sdX (in the server, it is
```

- Changer /etc/rear/local.conf et y ajouter ceci :

```
OUTPUT=USB
BACKUP=NETFS
BACKUP_URL="usb:///dev/disk/by-label/REAR-000"
```

- Exécuter ceci pour voir la configuration actuelle :

```
sudo rear dump
```

- Créer un disque de récupération (unmount REAR-000, puis ...)

```
sudo rear -v mkrescue # Just a rescue disk
sudo rear -v mkbackup # Rescue disk AND backup the system
sudo rear -v mkbackuponly # Only backup the system
```

- Pour restaurer le système, on boote depuis la clé USB. Lorsque le login est demandé, entrer `root`, puis `rear recover`. Redémarrer une fois la restauration terminée. Voir aussi <http://relax-and-recover.org/usage/>
- Commandes utiles :

```
# Changer la résolution d'écran
xrandr -fb 1920x1080 (-display:0)

# Inspection du système
inxi -GCS # General system info
top (-i) # hit q to quit
htop
# Boot and BIOS info
sudo dmidecode -t system
sudo dmidecode -t bios
uname -a # Basic system info
# Detailed hardware info (including eno1 network (1Gb Intel))
sudo lshw (-short)
sudo lshw -html > lshw.html
lscpu # CPU info
watch -n 2 sensors # CPU temperature
lsblk # Storage info
df -h
sudo disk -l
lsusb (-v) # Info about USB devices
# Reconfigure video driver
sudo update-initramfs -U
sudo dpkg reconfigure server-xorg
lsmod | grep nouveau
# Reconfigure locale
localectl US-utf8
# Diagnostic script for RStudio Connect
sudo /opt/rstudio-connect/scripts/run-diagnostics.sh /path/to/output/directory
```

- Installation du serveur MongoDB, voir <https://docs.mongodb.com/manual/tutorial/install-mongodb-on-ubuntu/> :

```
sudo apt-get install gnupg
wget -qO - https://www.mongodb.org/static/pgp/server-4.4.asc | sudo apt-key add -
echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu bionic/mongodb-org/4.4 multiarch" | sudo tee /etc/apt/sources.list.d/mongodb-org-4.4.list
sudo apt-get update
# Either latest version
sudo apt-get install -y mongodb-org
# ... or specific version
#sudo apt-get install -y mongodb-org=4.4.0 mongodb-org-server=4.4.0 mongodb-org-shell=4.4.0 mongodb-org-tools=4.4.0
# Optionally freeze version
#echo "mongodb-org hold" | sudo dpkg --set-selections
#echo "mongodb-org-server hold" | sudo dpkg --set-selections
#echo "mongodb-org-shell hold" | sudo dpkg --set-selections
#echo "mongodb-org-mongos hold" | sudo dpkg --set-selections
#echo "mongodb-org-tools hold" | sudo dpkg --set-selections
```

```

# Determine if it is systemd or init that is used
ps --no-headers -o comm 1 # systemd here
# Start mongodb
sudo systemctl start mongod

# By default, MongoDB only accept clients from same machine (what we want)
# Data directory is /var/lib/mongodb, but we want /data1/mongodb
sudo mkdir /data1/mongodb
sudo chown -R mongodb:mongodb /data1/mongodb
# Edit /etc/mongod.conf and adjust dbPath to /data1/mongodb
sudo nano /etc/mongod.conf

# Start the MongoDB server and test it from R
sudo systemctl start mongod
# If an error, first try sudo systemctl daemon-reload
# Status (can also use start/stop/restart
sudo systemctl status mongod
R
library(mongolite)
con <- mongo(collection = "shiny", db = "sdd")
con$insert(trees)
trees2 <- con$find()
all.equal(trees, trees2)
con$drop()
rm(con)
gc()
q("no")

# Quand c'est bon, activer le service définitivement avec
sudo systemctl enable mongod

# Create a program supervisor for common environment variables (not done yet)
mkdir -p /opt/scripts
touch /opt/scripts/connect-env.sh
nano /opt/scripts/connect-env.sh

# Add this in the file:

#!/bin/bash

echo arguments: "$@" >&2
echo >&2
export MONGO_URL="mongodb://sdd:sdd@sdd-umons-shard-00-00-umnnw.mongodb.net:27017,sdd-umons-sha
export MONGO_BASE="sdd"
exec "$@"

Then:
chmod 755 /opt/scripts/connect-env.sh

Add the following in /etc/rstudio-connect/rstudio-connect.gcfg
[Applications]
Supervisor = /opt/scripts/connect-env.sh

```

- Après un an, il faut mettre à jour le certificat de sécurité :

```
# Place sdd_umons_ac_be.cer|key in /etc/ssl/private. It seems both keys do not match:
sudo openssl x509 -noout -modulus -in /etc/ssl/private/sdd_umons_ac_be.cer | openssl md5
sudo openssl rsa -noout -modulus -in /etc/ssl/private/sdd_umons_ac_be.key | openssl md5
# Maybe the PKCS#7 format is correct? So, the .crt with that format can be converted this way:
sudo openssl pkcs7 -print_certs -in /etc/ssl/private/sdd_umons_ac_be.crt -out /etc/ssl/private/
sudo chmod o-r /etc/ssl/private/sdd_umons_ac_be.pem
sudo openssl x509 -noout -modulus -in /etc/ssl/private/sdd_umons_ac_be.pem | openssl md5
# Now, this is correct!

# Update the certificate: I receive a .zip file with both sdd_umons_ac_be.key and sdd_umons_ac_be.cer
# These files must be placed into /etc/ssl/private on the sdd server:
# From MacOS
scp ~/Desktop/sdd_certs_2021.zip econum@sdd.umons.ac.be:/data1/dump/cert.zip
# Create .pem file that invert the order of the 4 certificates from the .cer file, then...
scp ~/Desktop/sdd_umons_ac_be.pem econum@sdd.umons.ac.be:/data1/dump/sdd_umons_ac_be.pem
# From sdd
sudo ls -l /etc/ssl/private
sudo mv /etc/ssl/private/sdd_umons_ac_be.cer /etc/ssl/private/sdd_umons_ac_be.cer.old
sudo mv /etc/ssl/private/sdd_umons_ac_be.key /etc/ssl/private/sdd_umons_ac_be.key.old
unzip /data1/dump/cert.zip
rm /data1/dump/cert.zip
sudo mv sdd_umons_ac_be.* /etc/ssl/private
sudo chown root:root /etc/ssl/private/sdd_umons_ac_be.cer
sudo chown root:root /etc/ssl/private/sdd_umons_ac_be.key
sudo chmod 640 /etc/ssl/private/sdd_umons_ac_be.cer
sudo chmod 640 /etc/ssl/private/sdd_umons_ac_be.key
sudo mv /data1/dump/sdd_umons_ac_be.pem /etc/ssl/private/sdd_umons_ac_be.pem
sudo chown root:root /etc/ssl/private/sdd_umons_ac_be.pem
sudo chmod 640 /etc/ssl/private/sdd_umons_ac_be.pem
```

- J'ai un problème avec le driver vidéo... voici ce que j'ai dû faire pour le résoudre :

```
# Reconfigure the video driver
sudo update-initramfs -U
sudo dpkg reconfigure server-xorg
lsmod | grep nouveau
# Completely reconfigure the video driver if it gets stuck with Nvidia driver:
# Start in recovery mode: type <shift> repeatedly while booting
(sudo) update-initramfs -u

# Solved temporarily => could boot in X11, then, remove Nvidia drivers completely.
sudo apt-get purge nvidia-*
#sudo mv /etc/X11/xorg.conf /etc/X11/xorg.conf.bak
sudo apt-get install --reinstall xserver-xorg-video-intel libgl1-mesa-glx libgl1-mesa-dri xserver-xorg

# May not have any alternatives configured (ok to skip)
sudo update-alternatives --remove gl_conf /usr/lib/nvidia-current/ld.so.conf
# Does not change... still no keyboard
```

```

# Inactivate graph ics drivers repositories then ...
sudo dpkg -P $(dpkg -l | grep nvidia-driver | awk '{print $2}')
sudo apt autoremove
sudo apt install xserver-xorg-video-nouveau

    • Reconfiguration de la localisation

'localectl US-utf8

    • Mise à jour de R et RStudio Connect pour svbox2021 et nouvelle licence, 2021-05-31 :

# Current RStudio Connect is 1.8.2.1-12
ls /opt/R # See R versions already installed (3.4.4, 3.5.3, 3.6.3)
export R_VERSION=4.0.5
curl -O https://cdn.rstudio.com/r/ubuntu-1804/pkgs/r-${R_VERSION}_1_amd64.deb
sudo gdebi r-${R_VERSION}_1_amd64.deb
/opt/R/${R_VERSION}/bin/R --version # verification
sudo ln -s /opt/R/${R_VERSION}/bin/R /usr/local/bin/R
sudo ln -s /opt/R/${R_VERSION}/bin/Rscript /usr/local/bin/Rscript
# Install R package
sudo R
repos <-getOption("repos")
repos["CRAN"] <- "https://mran.microsoft.com/snapshot/2021-05-17"
options(repos = repos)
options(timeout = 300)

update.packages()

install.packages(c("tidyverse", "shiny", "bookdown", "data.table", "glue",
"here", "mongolite", "keras", "pastecs", "mlearning", "usethis", "testthat",
"covr", "sessioninfo", "reticulate", "remotes", "devtools", "knitr",
"latticeExtra", "inline", "Hmisc", "gridExtra", "gridGraphics", "ggsci",
"ggpubr", "GGally", "ggplotify", "ggrepel", "cowplot", "fs", "forcats", "purrr",
"R6", "RColorBrewer", "Rcpp", "anytime", "zoo", "assertthat", "bench", "hms",
"lubridate", "rsconnect", "RSQLite", "sos", "styler", "vctrs", "viridis",
"viridisLite", "withr", "xts", "igraph", "pryr", "proto", "renv", "tsibble",
"SciViews", "svMisc", "svGUI", "svDialogs", "extraDistr", "SuppDists", "lobstr",
"import", "miniUI", "vegan", "shinydashboard"))

install.packages("BiocManager")
BiocManager::install(c("graph", "ComplexHeatmap", "Rgraphviz",
"RDRTToolbox"), update = FALSE, ask = FALSE)

remotes::install_github("SciViews/mlearning@v1.0.6", force = TRUE)
# Fails! remotes::install_github("SciViews/tcltk2@v1.3.0", force = TRUE)
remotes::install_github("SciViews/svMisc@v1.2.0", force = TRUE)
remotes::install_github("SciViews/svGUI@v1.0.1", force = TRUE)
remotes::install_github("SciViews/svDialogs@v1.0.3", force = TRUE)
remotes::install_github("SciViews/svSweave@v1.0", force = TRUE)

```

```

remotes::install_github("SciViews/flow@v1.1.0", force = TRUE)
remotes::install_github("SciViews/data.io@v1.3.0", force = TRUE)
remotes::install_github("SciViews/chart@v1.3", force = TRUE)
remotes::install_github("SciViews/SciViews@v1.1.1", force = TRUE)

remotes::install_github("phgrosjean/pastecs@v1.4.0", force = TRUE)
remotes::install_github("phgrosjean/aurelhy@v1.0.8", force = TRUE)

remotes::install_github("rstudio/gradethis@ced5541")
remotes::install_github("r-lib/itdepends@f8d012b")

# Warning in svbox2021, we don't use learntdown anymore, but learnitdown instead!
remotes::install_github("SciViews/learnitdown@v1.3.0")
#remotes::install_github("BioDataScience-Course/BioDataScience@...")
#remotes::install_github("BioDataScience-Course/BioDataScience1@...")
#remotes::install_github("BioDataScience-Course/BioDataScience2@...")
#remotes::install_github("BioDataScience-Course/BioDataScience3@...")

# Check and update license
sudo /opt/rstudio-connect/bin/license-manager status
# (re)activate license
sudo /opt/rstudio-connect/bin/license-manager deactivate
sudo /opt/rstudio-connect/bin/license-manager activate YAG3-B2H5-MXQB-6HJS-E7WH-4G9Y-JITA
sudo systemctl restart rstudio-connect

# Upgrade RStudio Connect to version
curl -Lo rsc-installer.sh https://cdn.rstudio.com/connect/installer/installer-v1.9.1.sh
sudo -E bash ./rsc-installer.sh 1.8.8.2

```

- Installation de Python 3.8.5 :

```

sudo mkdir /opt/python
sudo curl -fsSL -o /opt/python/miniconda.sh https://repo.anaconda.com/miniconda/Miniconda3-lates
sudo chmod 755 /opt/python/miniconda.sh
sudo /opt/python/miniconda.sh -b -p /opt/python/miniconda

export PYTHON_VERSION="3.8.5"

sudo /opt/python/miniconda/bin/conda create --quiet --yes \
    --prefix /opt/python/"${PYTHON_VERSION}" \
    --channel conda-forge \
    python="${PYTHON_VERSION}" "pip<20.1"

# Check
/opt/python/"${PYTHON_VERSION}"/bin/python --version

```

- Installation de packages Python supplémentaires :

```

sudo /opt/python/"${PYTHON_VERSION}"/bin/pip install altair beautifulsoup4 cloudpickle \
    cython dask gensim keras matplotlib nltk numpy pandas pillow pyarrow \
    requests scipy scikit-image scikit-learn scrapy seaborn spacy sqlalchemy \

```

```

statsmodels tensorflow xgboost

# Add python to the system path (optional)
# add PATH=/opt/python/"${PYTHON_VERSION}"/bin:$PATH in /etc/profile.d/python.sh

# Make Python available as Jupyter Kernel (optional)
#sudo /opt/python/${PYTHON_VERSION}/bin/pip install ipykernel
#sudo /opt/python/${PYTHON_VERSION}/bin/python -m ipykernel install --name py${PYTHON_VERSION} -f

# Make Python available to reticulate
nano ~/.Rprofile # Add Sys.setenv(RETICULATE_PYTHON = "/opt/python/3.8.5/bin/python")

• Résoudre le problème de Pandoc: loadlocale.c:130: _nl_intern_locale_data:
  Assertion cnt < (sizeof (_nl_value_type_LC_TIME) / sizeof (_nl_value_type_LC_TIME[0]))' failed:

sudo mkdir -p /opt/scripts
sudo nano /opt/scripts/rstudio-connect-env.sh
# Add this:

#!/bin/bash
echo arguments: "$@" >&2
echo >&2
export LC_ALL=C
exec "$@"

# Then:
sudo chmod 755 /opt/scripts/rstudio-connect-env.sh

# Then, change RSconnect config:
sudo nano /etc/rstudio-connect/rstudio-connect.gcfg
# In this section, change Supervisor=

[Applications]
Supervisor = /opt/scripts/rstudio-connect-env.sh

# Then, restart RStudio connect
sudo systemctl restart rstudio-connect
# This does not work => reedit rstudio-connect.gcfg and comment out Supervisor= with ; and restart

# Need to rebuild the locales
# In /etc/locale.gen, I have only one item: en_US.UTF-8 UTF-8
# Then:
sudo mv /usr/lib/locale/locale-archive /usr/lib/locale/locale-archive.save
sudo locale-gen --no-archive
sudo locale-gen --no-archive en_US.UTF8

• Configuration de RStudio Connect pour le rendre sûre dans le cadre de tutoriels learnrs :

sudo apt-get install -y libapparmor-dev apparmor-utils

```

```
sudo R
install.packages("RAppArmor")
```

- Profil AppArmor :

```
#cd /usr/local/lib/R/site-library/RAppArmor/
cd /opt/R/4.0.5/lib/R/library/RAppArmor/
sudo cp -Rf profiles/debian/* /etc/apparmor.d/

# Verify the profile
sudo nano /etc/apparmor.d/rapparmor.d/r-user
# Make sure there is /usr/lib/rstudio/bin/pandoc/* rix,

#Load the profiles into the kernel
sudo service apparmor restart

#To disable enforcing the global R profile
sudo aa-disable usr.bin.r

#To start enforcing the standard R policy:
sudo aa-enforce usr.bin.r

# Set options for learnr tutorials
sudo nano /opt/R/4.0.5/lib/R/etc/Rprofile.site
```

- Ajouter ceci :

```
options(tutorial.exercise.evaluator.onstart = function(pid) {
  # Import RAppArmor
  require(RAppArmor, quietly = TRUE)
  # Set process group to pid (allows kill of entire subtree in cleanup)
  setpgid();
  # Set nice priority
  setpriority(10)
  # Set rlimits as appropriate
  rlimit_nproc(1000)
  rlimit_as(100*1024*1024)
  # Change to r-user profile
  aa_change_profile("r-user")
})

options(tutorial.exercise.evaluator.oncleanup = function(pid) {
  # Import RAppArmor
  require(RAppArmor, quietly = TRUE)
  # Kill entire process subtree. note that the second call works
  # because the call to setpgid above sets our pgid (process group id)
  # to our pid (process id)
  kill(pid, tools::SIGKILL)
  kill(-1 * pid, tools::SIGKILL)
})
```

- Le fichier r-user profile contient :

```

## Do not modify this file. Changes may be undone during upgrades.
## To play around, create a copy with a different profile name and
## put it in: /etc/apparmor.d/rapparmor.d/

profile r-user {
    #include <abstractions/base>
    #include <abstractions/nameservice>

    capability kill,
    capability setgid,
    capability net_bind_service,
    capability sys_tty_config,

    @{HOME}/ r,
    @{HOME}/.Rprofile r,
    @{HOME}/R/ r,
    @{HOME}/R/** rw,
    @{HOME}/R/{i686,x86_64}-pc-linux-gnu-library/** mrwix,

    @{PROC}/[0-9]*/attr/current r,

    /bin/* rix,
    /dev/tty r,
    /etc/R/ r,
    /etc/R/* r,
    /etc/fonts/** mr,
    /etc/resolv.conf r,
    /etc/xml/* r,
    /tmp/** mrwix,
    /usr/bin/* rix,
    /usr/include/** r,
    /usr/lib/gcc/** rix,
    /usr/lib/R/bin/* rix,
    /usr/lib{,32,64}/** mr,
    /usr/lib{,32,64}/R/bin/exec/R rix,
    /usr/local/lib/R/** mr,
    /usr/local/lib/R/site-library/** mrwix,
    /usr/local/share/** mr,
    /usr/share/** mr,
    /usr/share/ca-certificates/** r,
    /opt/rstudio-connect/ext/pandoc/** mrix,
    /opt/rstudio-connect/ext/pandoc/2.9/* mrix,
    /opt/rstudio-connect/ext/pandoc/2.11/* mrix,
    /opt/rstudio-connect/** mrwix,
    /opt/R/4.0.5/lib/R/** mr,
    /opt/R/4.0.5/lib/R/library/** mrwix,
    /opt/rstudio-connect/mnt/** mrwix,
    /opt/rstudio-connect/mnt/app/packrat/lib/x86_64-pc-linux-gnu/4.0.5/** mrwix,
}

```

- Redémarrer avec le profil modifié :

```
sudo nano /etc/apparmor.d/rapparmor.d/r-user
sudo service apparmor restart
```

- Note : ceci doit être fait sur le PC où les learnrs sont développés : `install.packages(c("RAppArmor", "unix"))`
- Utilisation d'une version différente de Pandoc (v 2.11 nécessaire à rstudio-connect) :

```
cd ~/Downloads
wget https://github.com/jgm/pandoc/releases/download/2.11.4/pandoc-2.11.4-1-amd64.deb
sudo dpkg -i pandoc-2.11.4-1-amd64.deb
pandoc --version # Still version 1.17.2
which pandoc # /usr/local/bin/pandoc
/usr/bin/pandoc --version # 2.11.4
/opt/rstudio-connect/ext/pandoc/2.11/pandoc --version # 2.11.2 with same options
/usr/lib/rstudio/bin/pandoc/pandoc --version # Same 1.17.2 than pandoc --version

#Sys.setenv(PATH = paste("/opt/rstudio-connect/ext/pandoc/2.11", Sys.getenv("PATH"), sep = ":"))
#Sys.setenv(PATH = paste("/usr/bin", Sys.getenv("PATH"), sep = ":"))
res <- system("which pandoc", intern = TRUE, ignore.stderr = FALSE, ignore.stdout = FALSE)
#stop(res)
#stop(Sys.getenv("RSTUDIO_PANDOC"))
#stop(rmarkdown::pandoc_version())
stop(rmarkdown::pandoc_available(error = TRUE))
(v2 <- ex$month[!duplicated(ex$month)])

# May be problem with pandoc-citeproc? But apparently not needed any more
which pandoc-citeproc # /usr/local/bin/pandoc-citeproc
pandoc-citeproc --version # 0.10.1

# pandoc-citeproc is missing from version 2.11... copy version from 2.9!
sudo cp /opt/rstudio-connect/ext/pandoc/2.9/pandoc-citeproc /opt/rstudio-connect/ext/pandoc/2.11/pandoc-citeproc

sudo cp /opt/rstudio-connect/ext/pandoc/2.11/pandoc /opt/rstudio-connect/ext/pandoc/2.11/pandoc
```

- Récupération des données depuis MongoDB ATLAS et backup des données (**note : remplacer d'abord par le mot de passe/ !!!**) :

```
echo $'#!/bin/bash
# Recuperation of MongoDB ATLAS data into local MongoDB database
# must be executed as root
# Copyright (c) 2021, Philippe Grosjean (phgrosjean@sciviews.org)

# First check that major versions of MongoDB servers (ATLAS vs local, x.y)
# match (otherwise, we may get corrupted data!)
mongo_local_version=`mongo --quiet --eval "db.version()" | awk -F "." '\{print $1 "." $2}\''`^
mongo_atlas_version=`mongo --quiet --eval "db.version()" "mongodb://sdd-teacher:<PASSWORD>@sdd-teacher:27017" | awk -F "." '\{print $1 "." $2}\'`^
if [ "$mongo_local_version" != "$mongo_atlas_version" ]; then
    echo "MongoDB versions mismatch: local=$mongo_local_version, ATLAS=$mongo_atlas_version" > /data/mongo_versions_mismatch.log`
```

```

    exit 0 # Use a higher value to report the error
else
    rm -f /data1/VERSIONS_MISMATCH
fi

# Make sure /data1/dump directory exists
cd /data1 &&
mkdir -p /data1/dump &&
chown econum:econum /data1/dump &&
# Record current date (in UTC format)
date -u "+%Y-%m-%d %H:%M:%S" > cur_date &&
# Is it a last_date file?
if [ -f "last_date" ]; then
    # Last date file exists, use it and get only data from that date
    # Need to specify collection here => call 3x for h5p, shiny & learnr
    mongodump --quiet --uri="mongodb://sdd-teacher:<PASSWORD>@sdd-umons-shard-00-00-umnnw.mongodb
    mongodump --quiet --uri="mongodb://sdd-teacher:<PASSWORD>@sdd-umons-shard-00-00-umnnw.mongodb
    mongodump --quiet --uri="mongodb://sdd-teacher:<PASSWORD>@sdd-umons-shard-00-00-umnnw.mongodb
else
    # No last_date file, get everything from MongoDB ATLAS
    mongodump --quiet --uri="mongodb://sdd-teacher:<PASSWORD>@sdd-umons-shard-00-00-umnnw.mongodb
fi

# If something is collected, an sdd subdirectory is created
if [ -d "/data1/dump/sdd" ]; then
    # Backup last_restore
    if [ -f "last_restore" ]; then
        cp last_restore forelast_restore
    fi
    # Backup last_date too
    if [ -f "last_date" ]; then
        cp last_date forelast_date
    fi
    mongorestore &> last_restore &&
    cp cur_date last_date
fi

# Clean up /data1/dump directory and go back to intial dir
rm -rf /data1/dump/* &&
cd - > /dev/null
' | sudo tee /usr/local/bin/get_atlas_data > /dev/null &&
sudo chmod 755 /usr/local/bin/get_atlas_data &&
sudo chmod +x /usr/local/bin/get_atlas_data

```

- Réalise un backup quotidien roulant, ainsi qu'un backup permanent chaque samedi :

```

# Regular daily backup
echo '#!/bin/bash
# Daily, temporary MongoDB database backup (overwritten next same day)
DIR=`date +%a` &&
DEST=/data1/backup/mongodb/$DIR &&
rm -rf $DEST &&

```

```

mkdir -p $DEST &&
mongodump -o $DEST &&
date -u "+%Y-%m-%d %H:%M:%S" > $DEST/backup_date
' | sudo tee /usr/local/bin/backup_data_daily > /dev/null &&
sudo chmod 755 /usr/local/bin/backup_data_daily &&
sudo chmod +x /usr/local/bin/backup_data_daily &&

# Weekly, more permanent backup
echo '#!/bin/bash
# Weekly, permanent MongoDB database backup
DIR=`date +%Y%m%d` &&
DEST=/data1/backup/mongodb/$DIR &&
mkdir -p $DEST &&
mongodump -o $DEST
' | sudo tee /usr/local/bin/backup_data_weekly > /dev/null &&
sudo chmod 755 /usr/local/bin/backup_data_weekly &&
sudo chmod +x /usr/local/bin/backup_data_weekly

```

Vérifier ces scripts et ensuite, ajouter une entrée crontab :

```
sudo crontab -e
```

Éditer le fichier en y rajoutant :

```

# Get data from MongoDB ATLAS every 5 min
*/5 * * * * /usr/local/bin/get_atlas_data
# Sync /data1/ with /data2/ every hour:47
47 * * * * /usr/bin/rsync -av --delete /data1/ /data2/ &> /data/datasync.log
# Create a daily (except saturday) and a weekly backup of the MongoDB database
# on 03:41:00
41 3 * * 6 /usr/local/bin/backup_data_weekly
41 3 * * 0,1,2,3,4,5 /usr/local/bin/backup_data_daily

```

- Ajout de capteurs matériels (contrôle de la température) :

```

sensors
sudo apt install nvme-cli
sudo nvme list
sudo nvme smart-log /dev/nvme0
sudo nvme smart-log /dev/nvme0 | grep temperature

```

Chapter 14

Rapport de progression

Un des gros problèmes des cours en classes inversées est d'arriver à **stimuler** les étudiants à faire le travail chez eux. Livrés à eux-mêmes, seulement un pourcentage trop faible d'entre eux va réaliser le travail demandé à domicile. Évidemment, c'est également préjudiciable au travail en classe, si ce dernier part du principe que la matière hors présentiel est censée être vue. Et si des travaux de groupe (conseillé) sont prévus, ceux qui n'ont rien fait apparaîtront comme des boulets pour les autres.



En parallèle au déploiement du matériel d'e-learning hors présentiel, nous devons mettre également en place des outils pour permettre à l'étudiant de suivre sa progression d'apprentissage, et aussi pour le stimuler dans cet auto-apprentissage.

Les outils qui peuvent être mis en place sont :

- *Pénalisation si le travail n'est pas fait.* Il faut arriver à déterminer ici objectivement si le travail est fait ou pas. Le plus simple est d'interroger les étudiants en début de chaque classe en présentiel, **mais cela prend du temps et n'est pas très constructif** ! Cette approche n'est donc pas conseillée.

La vidéo suivante explique d'ailleurs pourquoi le coercitif ne fonctionne pas bien et pourquoi il faut plutôt **valoriser le travail effectué** plutôt que de punir s'il n'est pas fait. Il faut aussi **une récompense immédiate**, et favoriser l'**émulation individuelle** (se dépasser soi-même), et **de groupe** (se comparer à la progression générale de la classe).

Les aides et encouragements au travail que nous avons mis en place de manière utile sont :

- L'*émulation personnelle* au travers de points reçus si l'on fait tous les exercices (les badges, et la “gamification” en général de l'apprentissage sont également à considérer sérieusement dans ce contexte),
- La *valorisation d'un comportement ou d'un résultat positif au delà de la moyenne* à l'aide de points bonus (on est toujours dans la gratification et la gamification ici),
- Montrer que nous (en qualité d'enseignants), nous nous *intéressons réellement à leur travail*. Cela tient à des petits détails : des retours sous forme de courts commentaires, des relances, des questions, des encouragements, la reconnaissance de leurs progrès, ...
- l'*entre-aide*, faire expliquer les points difficiles par un collègue qui les ont compris est extrêmement efficace. En plus, cela permet de mitiger un aspect difficile à gérer : la différence de vitesse d'apprentissage entre les “bons” et les “mauvais” étudiants (non, il n'y a pas de bons et de mauvais étudiants, nous le savons, il y a juste des étudiants avec des facultés et des motivations différentes... mais certains sont quand même beaucoup plus lents et **énervants** que d'autres). Au lieu de se tourner les pouces, demander à ceux qui ont fini en avance d'expliquer aux autres permet d'utiliser efficacement le temps et les ressources humaines en classe.
- Les *travaux de groupes*, les *challenges* (par exemple, les étudiants doivent reproduire un graphique difficile), les *compétitions* (une course contre le temps et/ou entre plusieurs groupes pour résoudre un problème) participent efficacement à dynamiser le travail en classe et à varier les activités.

- L'utilisation des *réseaux sociaux* pour communiquer. Le forum de Moodle est peu ou pas utilisé du tout. Les issues de GitHub, une fois que les étudiants ont appris à les utiliser, s'avèrent utiles car elles sont disponibles près du dépôt lui-même qui contient les exercices. Mais ce qui fonctionne le mieux (testé en période de confinement Covid-19), c'est *Discord* parce que beaucoup d'étudiants connaissent et utilisent déjà pour les jeux. Le ton des échanges y est d'ailleurs plus décontracté (parfois nettement plus), mais cela reste un outil efficace pour aider un étudiant qui est bloqué dans sa progression.
- Les étudiants sollicitant une *aide plus directe* apprécient de pouvoir poser leurs questions par email. Afin de pouvoir être plus réactif, nous avons mis en place une adresse `sdd@sciviews.org` pour laquelle tous les enseignants du cours ont accès. Ainsi, les interventions peuvent être partagées et plus rapides.
- Pour obtenir un *retour plus direct* par rapport aux activités, nous mettons progressivement en place le **rapport de progression**. Les actions de l'étudiant dans les learnrs (et dans un second temps également H5P et les apps Shiny) sont enregistrées dans une base de données MongoDB. Le rapport de progression est une app Shiny qui collecte, nettoie, prépare, et présente graphiquement la progression de l'étudiant. L'*émulation de groupe* est obtenue en comparant la progression de l'étudiant à la progression générale de la classe.

Afin d'avoir un retour immédiat, nous souhaitons rendre ce rapport de progression dynamique de sorte qu'il montre l'évolution en temps réel. Il est en effet important que l'étudiant puisse visualiser sur son rapport qu'il vient d'avancer en achevant à la seconde un exercice ! De plus, nous souhaitons associer cela à des récompenses sous forme de **badges** lorsque l'étudiant franchit une étape dans son apprentissage (aussi, des badges spéciaux pourraient être associés aux points bonus). La gestion des badges a été abordée dans la section dédiée à Moodle (voir chapitre 12).

14.1 Présentation générale

L'app Shiny de rapport de progression est développée dans le dépôt GitHub privé `sdd_progress-report`. Comme il contient du code sensible, il n'est pas disponible de manière publique. Par contre, les enseignants du cours y ont accès. Le fichier `README.md` du dépôt détaille son fonctionnement et contient du code pour le tester.

Cette app Shiny est ensuite intégrée dans Moodle via un lien URL qui ouvre l'app dans une fenêtre interne et lui passe des paramètres issus de Moodle (identité de l'étudiant, cours correspondant) afin de déterminer quel rapport il faut présenter.

14.2 Les données

Nous utilisons MongoDB comme système de stockage de nos données. Ce type de base de données permet de stocker des données semi-structurées qui correspondent bien à ce que les événements xAPI renseignent. Par sécurité, une base de données MongoDB *locale* est créée sur le serveur RStudio Connect. Elle n'est pas disponible depuis l'extérieur. Donc, seules les applications exécutées à partir de RStudio Connect y ont accès. Par contre, cela ne permet pas d'enregistrer les données générées ailleurs (H5P, applications learnr ou shiny exécutées dans la SciViews box, ...). Ainsi, nous déployons une seconde

base MongoDB sur le cloud via MongoDB ATLAS. Cette dernière sert uniquement à collecter *transitoirement* les données externes au serveur. Ensuite, un script sur RStudio Connect est exécuté à intervalle régulier pour rapatrier ces données externes dans la base de données interne.

Une autre solution aurait été d'ouvrir la base de données associée à notre serveur RStudio Connect, mais la solution à deux points d'accès offre les avantages suivants :

- Sécurité maximale des données une fois rapatriées dans la base sur le serveur
- Duplication des points d'accès (et MongoDB ATLAS utilise 3 serveurs différents sur le cloud) afin de garantir que les données puissent toujours être récoltées, même si le serveur RStudio Connect ne répond plus

14.2.1 Base interne

TODO: expliquer l'installation et le fonctionnement de la base MongoDB interne.

A des fins de test et développement, la base de donnée interne MongoDB peut être ouverte transitoirement. Sur le serveur RStudio Connect, le port de MongoDB 27017 est ouvert dans les deux sens. Ainsi, il est possible de se connecter depuis RStudio Connect vers MongoDB ATLAS pour en récupérer les données d'une part, et il est aussi possible de modifier temporairement le fichier de configuration de MongodB local pour permettre un accès externe lors de la mise au point des applications Shiny qui doivent y accéder.

Ce fichier de configuration est `/etc/mongodb.conf` et il contient. Pour modifier sa configuration transitoirement, nous pouvons nous y connecter via VPN UMONS (`ssh econum@sdd.umons.ac.be` avec mot de passe econum classique), et ensuite :

```
sudo nano /etc/mongodb.conf
# Editer la ligne bindp: 127.0.0.1 en bindp: 127.0.0.1,193.190.194.75
# Ensuite Ctrl-X + Y(es)
# Enfin, redémarrer le service MongoDB
sudo systemctl restart mongodb
```

Pour tester l'accès :

```
# Accès en local (il faut succeeded!)
nc -zv 127.0.0.1 27017
# Accès distant
nc -zv 193.190.194.75 27017
# En local pour déterminer l'état du port 27017
netstat -pano | grep 27017
```

On peut aussi directement tester l'accès à la base MongoDB avec les outils en ligne de commande installés depuis un ordinateur distant comme ceci :

```
mongo "mongodb://193.190.194.75:27017"
db
quit()
```

Pour refermer le port, on fait l'opération inverse : limiter `bindp= 127.0.0.1` seulement dans le fichier de configuration.

14.2.2 Base externe

TODO: expliquer l'installation et le fonctionnement de la base MongoDB externe. Deux comptes: **sdd** avec des droits limité uniquement en écriture pour l'introduction des données et **sdd-teacher** avec des droits en lecture et écriture.

14.2.3 Gestion des utilisateurs

Nous récupérons les informations provenant de Wordpress concernant les utilisateurs. Ensuite, le rapport de progression est affiché dans un iframe en fonction des données issues de Moodle, y compris **iref** qui est la référence interne à Moodle et qu'il est impossible de deviner par ailleurs (en effet, les autres données nom, email, cours, etc. pourraient être devinées et un tiers pourrait visualiser le rapport de progression d'un étudiant sans son consentement... sauf qu'il lui faut aussi **iref** qui joue donc garde-fou ici).

La récupération des données des utilisateurs est détaillée dans le dépôt privé **sdd_student_2020**. En 2020-2021, nous avons un bug qui fait que les données Moodle ne sont pas forcément enregistrées comme métadonnées dans Wordpress si l'utilisateur ne va pas sur Moodle *après* avoir entré son login GitHub. Ce problème devra être réglé plus tard, mais pour l'instant, nous avons demandé à tous les étudiants d'effectuer cette opération (il suffit de la faire une seule fois). Un problème subsiste cependant pour les étudiants en passerelle qui suivent deux cours, car l'entrée **icourse** ne peut contenir qu'une seule valeur. Nous devons donc résoudre ce problème avec du code spécifique R... mais ce sera à revoir pour une version ultérieure.

14.2.4 Format des données

A partir de 2020, avec l'apparition de {learnitdown}, nous diversifions l'offre d'exercices enregistrés. Préalablement, il s'agissait uniquement de learnr exécutés dans la SciViews Box. A présent, les H5P et les applications Shiny peuvent également être enregistrées.

Afin d'éviter la multiplication des formats de données, nous avons donc homogénéisé leur présentation entre les différentes applications dans {learnitdown}. Ces données sont toutefois enregistrées dans des collections différentes (**h5p**, **learnr** et **shiny**) dans la base **sdd**. Chaque entrée contient les champs suivants :

- **__id:** C'est l'identifiant unique du document MongoDB attribué automatiquement lors de l'insertion de chaque entrée
- **session:** La session à partir de laquelle l'évènement a été généré. Cette information est surtout utile pour les applications Shiny. Pour H5P, c'est un identifiant complémentaire de l'utilisateur enregistré qui est indiqué. Le plus souvent du type **email: mailto:user@site.com**, mais pas forcément. Pour l'instant, les learnr n'introduisent rien dans ce champ.
- **date:** La date au format GMT à laquelle l'évènement a été généré. Cette date est enregistrée à la microseconde près, mais la résolution est inférieure (probablement 50 à 150μsec, sauf pour les applications Shiny où le temps est comptabilisé à la milliseconde près).
- **id:** Uniquement pour H5P, l'identifiant du widget, le numéro à rentrer dans le shortcode, par exemple, `[h5p id="3"]` pour l'item 3.

- **app:** L'identifiant textuel de l'application, voir convention pour le nom des applications ci-dessous.
- **version:** Le numéro de version pour les apps Shiny et learnr. Les applications H5P n'ont malheureusement pas de numéro de version. Ce champ contient donc `null` dans ce cas, sauf s'il s'agit d'un "sous-contenu" (voir convention de noms d'apps ci-dessous), dans ce cas, nous aurons l'identifiant du sous-contenu.
- **user:** Utilisateur actuel sur la machine où l'application s'exécute.
- **login:** Login GitHub/Wordpress de l'utilisateur (peut être le même ou différent de **user**).
- **email:** Adresse mail institutionnelle (valeur passée par Moodle dans `iemail`). Pour les applications H5P, l'email Wordpress est resigné généralement dans session, s'il a été défini. Pour les applications Shiny, l'identification complète de l'utilisateur, y compris son email Wordpress se retrouve dans le champ **data** pour l'évènement **started**.
- **course:** Le cours que suit l'étudiant, sous forme de l'identifiant Moodle (ex.: `S-BIOG-006`).
- **institution:** Nom de l'institution où l'étudiant est enrôlé.
- **verb:** Le verbe xAPI correspondant à l'évènement (voir ci-dessous verbes xAPI).
- **correct:** "TRUE" ou "FALSE" uniquement pour les évènements correspondant à des réponses à une question, donc les verbes **answered** (pour les questions) ou **submitted** (pour du code R). Dans les autres cas, sa valeur est "", sauf si une application Shiny a été lancée, mais sans que l'utilisateur ne clique jamais sur le bouton **Submit**. Dans ce cas, la valeur vaut "NA".
- **score:** Nombre de points obtenus pour la question. Dépend de l'application. Pour les H5P, il s'agit du nombre d'items corrects moins le nombre d'items erronés. Pour Shiny, c'est le nombre de widgets correctement positionnés par rapport au nombre testé dans la solution. Enfin pour learnr, c'est 0 ou 1 selon que la question est complètement correcte ou pas. Pas pertinent en dehors de réponses et donc ce champ vaut `null` dans ce cas.
- **max:** Score maximum pouvant être obtenu pour la question. Dépend du contexte. En dehors de réponses aux questions, ce champ vaut 0.
- **grade:** La fraction de bonne réponse, ou autrement dit, une note sur 1 pour la question. Si on a 1 la réponse est totalement correcte, mais si on a 0.5, elle ne l'est qu'à moitié. En dehors de réponse à une question, ce champ n'a pas de sens et sa valeur est `null`.
- **label:** Le libellé de l'item générant l'évènement. Pour Shiny, c'est le nom du widget. Pour H5P, c'est le texte de la question, etc.
- **value:** La valeur sélectionnée pour le widget Shiny, ou la réponse donnée pour H5P, par exemple. Dépendant du contexte.
- **data:** Des données complémentaires sous forme d'une chaîne de caractères contenant un objet JSON. Pour H5P, il s'agit de l'évènement xAPI généré. Pour learnr, il s'agit d'une partie du contenu (moins **correct** et **label** qui sont extraits dans les champs correspondants). Pour Shiny il s'agit d'informations complémentaires sur les widgets.

14.2.4.1 Convention de nom des apps

Nous avons potentiellement beaucoup d'applications H5P/Shiny/learnr réparties un peu partout dans les différents cours. Pour s'y retrouver, nous utilisons donc une convention de noms telle que nous pouvons déterminer rapidement à quelle partie de quel cours l'app se rattache. Cette convention est la suivante :

- Les noms commencent par A00a_ avec :
 - La première lettre en majuscule correspondant au cours (A = SDD I, B= SDD II, etc.). Pour les applications non attribuées à des cours en particulier, nous utilisons . pour toutes les applications utilisées dans la présentation générale, et sinon, Z pour tout le reste.
 - Deux chiffres indiquant le module où se trouve l'app. On utilise 00 s'il n'y en a pas ou si elle apparaît dans le prambule du bookdown, ou 99 pour une app apparaissant dans les appendices.
 - Une lettre minuscule servant à ordonner les applications d'un même module. Ceci est surtout utile pour les learnrs afin qu'ils apparaissent dans le bon ordre dans l'onglet 'Tutorials' de RStudio. Pour les autres, c'est utilisable mais non nécessaire. On peut aussi utiliser le point . pour ne rien spécifier ici.
 - Le trait souligné qui sépare le code d'identification du nom de l'application.
- Après le code, le nom de l'application doit être court et informatif. Séparer les mots par des tirets.
- Dans le cas particuliers des "sous-contenus" de H5P (des widgets à l'intérieur d'autres widgets, exemple : une question à choix multiple à l'intérieur d'une présentation ou une vidéo), on utilisera le même code et nom que le widget principal, suivi du slash et du nom attribué au sous-contenu. Par exemple, la question 1 dans une présentation pourrait s'appeler B02a_présentation/question1 si la présentation est dans le cours SDD II (B), module 2 (02), première position a, et qu'elle se nomme elle-même présentation.

Le système de nommage des applications choisi est un peu compliqué, mais avec l'habitude, il permet d'identifier instantanément les données dans la base par rapport à leur contexte !

14.2.4.2 Verbes xAPI

Les évènements xAPI de H5P sont déjà décrits par des verbes. Par contre, les évènements Shiny ou learnr sont décrits via des dénominations propres. Afin d'homogénéiser le tout, nous avons décidé de traduire les évènements Shiny et learnr en verbes xAPI selon le tableau suivant :

Verbe xAPI	learnr	shiny	H5P	Remarque
started	session_start	start		L'utilisateur a démarré activement une application

Verbe xAPI	learnr	shiny	H5P	Remarque
attempted			attempted	L'utilisateur arrive sur un H5P (différent de started)
exited		inputs/quit		L'utilisateur quitte volontairement (différent de stopped)
stopped	session_stop	stop		L'application s'est arrêtée
displayed	section_viewed			L'utilisateur visualise une section
progressed	section_skipped		progressed	L'utilisateur avance dans les exercices (différent de displayed)
seeked	video_progress			Progression dans une vidéo
interacted		inputs	interacted	Dans Shiny: sauf boutons 'submit' & 'quit'
answered	question_submission		answered	Réponse à une question (hors code R)
reset	reset_question_submission			Nouvel essai après une réponse erronée
executed	exercise_submission			Bouton 'Run Code' dans learnr
evaluated	exercise_result			Résultat de l'évaluation du code R, si correct ==== " "

Verbe xAPI	learnr	shiny	H5P	Remarque
submitted	exercise_result	inputs/result		Idem, mais si correct != “ ” dans learnr, Bouton Shiny ‘submit’
computed		outputs		Par défaut, les outputs ne sont pas enregistrés dans Shiny
debugged		errors		Erreur détectée dans une app Shiny
assisted	exercice_hint			Sauf si dernier hint = réponse => revealed
revealed	exercice_hint			Pour les boutons ‘solution’ dans learnr & dernier ‘hint’
completed		completed	Dans H5P, un bilan général est calculé à la fin	

A noter que ces correspondances ont été établies sur base de la définition des verbes xAPI, ainsi que sur la liste des évènements learnrs.

Chapter 15

Dashboard

TODO: documentation about {learnitdashboard}.

Part III

More tools

Chapter 16

Vidéos

TODO: expliquer ici les outils pour la création de vidéos et l'intégration à la plateforme LearnIt::R.

Chapter 17

Wooclap

TODO: présenter Wooclap et comment l'intégrer dans la plateforme ici...

Appendix A

Installation

TODO: installation des outils logiciels nécessaires pour déployer la plateforme pédagogique...

Appendix B

Maintenance

Le système dans son ensemble est complexe et nécessite évidemment un minimum de maintenance. Les différentes opérations sont listées ici.

B.1 Maintenance régulière

TODO...

B.2 Nettoyage à la fin de chaque cours

En aout et en janvier, une maintenance supplémentaire est nécessaire pour préparer respectivement au premier quadrimestre de l'année académique suivante et au second quadrimestre de la même année. Les opérations suivantes sont à réaliser :

- Vérifier le bon fonctionnement et la mise à jour de toutes les composantes.
- Contrôle de la table des utilisateurs dans WordPress + dernière exportation et importation dans p-Cloud et dans la base de données MongoDB de sdd.umons.ac.be
- Élimination des comptes d'utilisateurs arrivés en fin de cursus
- Nettoyage et pesudonymisation des dépôts GitHub relatifs aux étudiants arrivés en fin de cursus dans l'espace BioDataScience-Course (TODO !!!)
- Backup de données H5PxApiKatchu éventuelles et nettoyage de ces données
- Backup complet du site WordPress avec UpdraftPlus
- Récupération d'un snapshot de WordPress dans p-Cloud
- Récupération d'un backup de la base de données MySQL dans p-Cloud. Aller dans phpMyAdmin, exporter au format .sql (quick) dans /srv/data/tmp. Ensuite, récupérer le fichier, le zipper et le placer sur p-Cloud pour archivage
- Update éventuel WordPress après ces backups (PHP et/ou MySQL)

- Vérification de la synchronisation entre MongoDB ATLAS et MongoDB sdd.umons.ac.be et nettoyage des tables h5p, shiny et learnr sur MongoDB ATLAS
- Récupération et nettoyages des backups MongoDB depuis sdd.umons.ac.be et réalisation d'un snapshot local archivé sur p-Cloud. Pour récupérer les données entre deux dates seulement :
`mongodump --query "{\"ts\":{\"$gt\":{\"date\":date -d 2011-08-10 +%s000},\"$lte\":{\"date\":d 2011-08-11 +%s000}}}"`
- Récupération d'une copie des différentes bases de données SQLite liées aux challenges, et élimination de ces bases de données depuis sdd.umons.ac.be à la fin de chaque année académique.
`sudo tar -czvf backup/challenges.tar.gz A04_challenge B04_challenge C03_challenge, then from a local machine : scp econum@sdd.umons.ac.be:/data1/backup/challenges.tar.gz challenges.tar.gz`
- Mise à jour et vérification logicielle et matérielle de sdd.umons.ac.be
- Changement de tous les mots de passe et génération de nouveaux fichiers cryptés avec les informations de connexion (en début d'année académique uniquement)
- Remise à zéro de tous les flags d'utilisateurs dans WordPress

Appendix C

Anonymisation des données

Les informations enregistrées via `{learnitdown}` dans des bases de données MongoDB servent à évaluer la progression des étudiants dans leur apprentissage, et à les noter à la fin du cours. Ces données peuvent également se révéler précieuses pour la recherche en pédagogie car elles contiennent des indices intéressants pour déterminer ce qui fonctionne bien ou moins bien. Cependant, les données contiennent des informations à caractère personnel, ce qui en limite le champ d'utilisation (en respect du RGPD en Europe, par exemple). Une diffusion de ces données à des fins de recherche en pédagogie n'est possible qu'après élimination des informations à caractère personnel, c'est-à-dire, après avoir **anonymisé** les données.

Pour que les données restent utiles, il faut toutefois continuer à attribuer les données à différents individus anonymes distincts. Ainsi, chaque donnée doit être attribuée à un pseudonyme type “student1”, “student2”, ..., “studentN” *de manière consistante à travers tout le jeu de données*. Cela ne pose pas de problèmes particuliers pour les données des exercices H5P/Shiny/learnr dans la base de données MongoDB. Il suffit de remplacer les logins des étudiants pour les pseudonymes choisis à l'aide d'une table de conversion (TODO: faire une fonction dans `{learnitdown}` pour cela et montrer ici un exemple de code). Par contre, c'est nettement plus compliqué en ce qui concerne les dépôts GitHub. Des outils existent pour anonymiser partiellement ou totalement des dépôts GitHub, mais ils ne sont pas faciles à mettre en œuvre ou présentent des limitations. TODO: il faudra ici trouver une solution qui convienne à notre cas particulier, par exemple : <https://docs.github.com/en/github/using-git/changing-author-info>, qui peut déjà être utile pour réattribuer des commits à un étudiant s'il a entré une mauvaise identité (souvent, c'est ‘sv’). Et un plugin git utilise cette approche pour complètement anonymiser un dépôt Git tout entier: <https://github.com/cvortmann/git-anonymize>. Malheureusement, si je comprend bien, tout le monde est modifié en “anonymous”, or il faudrait mettre “student1”, “student2”, etc., comme dans la base de données H5P/learnr/shiny. Un beau projet pour un étudiant en informatique, peut-être ?

Enfin, concernant les fils de discussion sur Discord, il est assez facile de les copier dans des documents Word, de les exporter dans un autre format (Markdown ?), et puis de les anonymiser à l'aide de code écrit en R. **Faire attention ici aux copies d'écrans que les étudiants joignent parfois à leurs posts et qui peuvent contenir des informations personnelles !**

Références